공학석사학위논문

Uncertainty Quantification of Reservoir Performances Using Streamline Based Inversion and Distance Based Method

유선기반 역산 및 거리기반 방법을 이용한 저류층 거동의 불확실성 정량화

2014년 2월

서울대학교 대학원

에너지시스템공학부

박 지 훈

Uncertainty Quantification of Reservoir Performances Using Streamline Based Inversion and Distance Based Method

유선기반 역산 및 거리기반 방법을 이용한 저류층 거동의 불확실성 정량화

지도교수 최 종 근

이 논문을 공학석사학위논문으로 제출함

2014년 2월

서울대학교 대학원

에너지시스템공학부

박 지 훈

박지훈의 석사학위논문을 인준함

2014년 2월

위 원 장 박 형 동 (인)

부 위 원 장 최 종 근 (인)

위 원 송 재 준 (인)

Abstract

For decision makings, it is crucial to have proper reservoir characterization and uncertainty assessment of reservoir performances. Since an initial model constructed with limited data has high uncertainty, it is essential to integrate both static and dynamic data for reliable prediction. Uncertainty quantification is computationally demanding because it requires a lot of iterative forward simulations and optimizations in a single history matching. Multiple realizations of reservoir models should be history matched. In addition, history matching is mathematically a highly ill-posed problem.

In this paper, a methodology is proposed to rapidly quantify uncertainties by combining streamline based inversion and distance based method. First, a distance between each model is defined as the norm of differences in generalized travel time vectors. Second, they are grouped according to distances and representative models are selected instead of matching all models. Third, generalized travel time inversion is applied for integration of dynamic data and a streamline simulator is adopted as a forward simulator to take advantage of computational efficiency. It is verified that the proposed method gathers models with similar dynamic responses and permeability distribution. It also assesses the uncertainty of reservoir performances fairly well, while reducing the amount of calculations significantly by using the representative models.

Keywords: streamline simulation, generalized travel time inversion, distance based method, reservoir characterization

Student Number: 2012-20999

Table of Contents

List of Tables

List of Figures

1. Introduction

An initial reservoir model has high uncertainty due to limited data available and lack of understanding of phenomena in the subsurface. Therefore, it is imperative to reconcile static and dynamic data obtained with the initial model to reliably predict future performances, estimate reserves, and assess the uncertainty. This overall process in reservoir engineering is called reservoir characterization, which is indispensable for decision making. For decision making, the probabilistic estimates which can be offered by uncertainty quantification, are often the most meaningful (Oliver et al., 2008).

Static data can be integrated via geostatistical techniques. Dynamic data, which are dependent on time, are integrated through inverse modeling or history matching. For uncertainty quantification, a reservoir engineer should overcome some challenges. First, uncertainty quantification requires excessive calculations. A myriad of forward simulations and optimizations should be carried out. In addition, multiple realizations of the reservoir are needed for history matching. Especially, when inverse modeling of a field-scale large reservoir is performed, enhancing computational speed is critical for feasible uncertainty quantification because a single forward simulation takes significant amount of calculations.

Second, history matching is mathematically a highly ill-posed problem. The reservoir is characterized by several model parameters, such as permeability, porosity, or specific geological features. However, available data, which are from cores, well logging, or seismic, are limited. Therefore, there exist multiple solutions that satisfy the given

production history. In addition, it would yield unstable solutions, i.e., the model parameter is sensitive to small perturbations on observed data.

Third, an inverse problem for flow in porous media is strongly nonlinear. If an initial model is far from the true field which we seek to find, an optimization algorithm might fail to spot the global minimum, and the solution would be stuck in local minima. This is intensified in sensitivity or gradient based optimization algorithms. Gradient-free methods such as genetic algorithm or simulated annealing theoretically ensure to spot a global minimum. However, they are computationally prohibitive, of which the characteristics are less suitable for the uncertainty quantification.

Fourth, streamline simulator can be an effective forward simulator to reduce the computational cost. Streamline simulation is a kind of reservoir simulation algorithms to model subsurface flow by describing 3-D flow as an integration of 1-D streamlines (Datta-Gupta, 2000; Jang and Choe, 2001; Cheng, 2005). The key principle of streamline simulation is that the governing partial differential equations are transformed to and solved on 1-D time of flight (TOF) coordinate instead of 3-D Cartesian coordinates. It renders the simulator computationally rapid compared to a conventional finite difference simulator. Therefore, a streamline simulator is an efficacious tool in solving large, geologically complex and heterogeneous systems (Thiele, 2005). With these attributes, streamline simulation has been widely used in reservoir engineering, such as swept volume calculations, rate allocation and pattern balancing, and production data integration (Datta-Gupta, 2000). In this research, streamline simulation is used for both forward simulations and inversions to take advantage of its computational speed.

A streamline based approach also offers powerful methodologies in inversion. Emanuel and Milliken (1998) took the streamline based approach in history matching.

They used streamlines to allocate gridblocks to wells to identify gridblocks that affect performances of production wells. The reservoir properties of assigned gridblocks are altered and simulations are repeated until the model is conditioned to the production history. They named their method as assisted history matching (AHM) because it is to assist manual history matching to designate gridblocks of which the properties should be adjusted. Milliken et al. (2001) successfully applied AHM to history match real reservoirs of which the number of cells range from 10^5 to 10^6. They demonstrated that the streamline based inversion has an advantage over other history matching techniques in a sense that the perturbation to match a certain well has less influence on the performances of other wells.

Vasco et al. (1998) proposed analytic techniques that offer the foundation of streamline based automatic history matching. Applying the analogy from seismic ray tracing to streamlines, they formulated parameter sensitivities in terms of an analytic integral along the streamline. They suggested two stage inversions. In detail, production responses are matched after breakthrough times from producing wells are matched. Using this approach, they integrated dynamic data with the model using inversion techniques from geophysical inverse theories. The matching of time is called travel time inversion. Matching production responses, which is a conventional way to process data in history matching, is called amplitude inversion. They applied their method to a large scale reservoir and verified that automatic history matching with streamline derived sensitivities successfully integrates the data rapidly.

Wu and Datta-Gupta (2001) proposed the concept of generalized travel time inversion (GTTI) and proved the feasibility for application to a large-scale field. In GTTI, the data misfit is defined as generalized travel time (GTT). GTT is an optimal

time shift that minimizes the square sum of amplitude differences. They demonstrated that it has an advantage over other inversion techniques in terms of computational costs because the sensitivity computation depends only on the number of wells. They also pointed out that if the data misfit is defined as time rather than the amplitude, the optimization result is relatively insensitive to the prior model because of quasilinearity. This was reconfirmed by Cheng et al. (2003). They quantified the nonlinearity measure, and proved that the nonlinearity measure is orders of magnitude lower in travel time inversion and GTTI than the amplitude inversion.

He et al. (2001) suggested how to calculate GTT sensitivities. They regarded that each dynamic response is shifted by the same amount of time, and averaged the travel time sensitivities. In addition, they derived the additional terms to add to travel time sensitivities to include pressure updates when the field conditions change.

Cheng et al. (2003) assessed the uncertainty by applying GTTI with randomized maximum likelihood (RML) method. They substantiated that GTTI can successfully quantify the uncertainty due to its high computational efficiency. They generated multiple realizations and matched them to the realizations of the observed data. They employed GTTI to minimize the objective function in RML.

Cheng et al. (2005) proposed the methodology to use a finite difference simulator with 3 phase GTTI. By using a finite difference simulator as a forward simulator, they could overcome shortcomings of streamline simulation. Even though there are additional calculations to generate streamlines and trace them, they hardly deteriorate the computational speed, because the amount of calculations is small compared to the total amount of calculations.

Schedit and Caers (2009) proposed how to assess the spatial uncertainty by taking the concept of distance function. The distance is a measure of similarity (or dissimilarity) between geological models. By modeling multidimensional space of uncertainty with kernel techniques, they gathered similar models within each cluster and selected representative models. They showed that a few selected models sufficiently capture the uncertainty of the reservoir models. They validated their method by applying to a deepwater turbidite offshore reservoir.

The limitation of previous studies are as follows. Even though AHM brought significant enhancement in history matching, it is still manual history matching, which involves subjective judgments. As Cheng et al. (2004) pointed out, it may result in 'tube like' artifacts into the geological model since it is assumed that a certain producing well is only affected by gridblocks that are assigned by streamlines. In addition, if there are a large number of wells, it is still time consuming and laborious. Therefore, to preserve the geological realism and avoid subjective bias, automatic history matching should be utilized.

In Cheng et al. (2003)'s work, one hundred models were generated and all of them were matched through RML and GTTI. However, if similar models are grouped and inversions are performed on representative models only, it will reduce the number of calculations significantly in order to assess the uncertainty. Furthermore, considering that the more realizations offer the more reliable uncertainties, it is imperative to reinforce the computational efficiency when assessing the uncertainty of a large field scale reservoir.

Schedit and Caers (2008, 2009) defined the distance as the flow based distance. The distance was defined as the differences of field oil production rates with respect to time

(Schedit and Caers, 2008) or squared differences of cumulative oil or water production (Schedit and Caers, 2009). Compared to static distance, it gives more accurate result for a specific field. However, they did not consider the effect of individual wells but the sum of the responses from each well. Applying the concept of GTT in the distance, it will take effects of individual wells since sensitivities are calculated along streamlines connecting the injection well and the production wells.

The main objective of this study is to propose a new methodology of uncertainty quantification by combining streamline based inversion and distance based method. First, a new concept of distance using GTT is proposed. Second, similar models are grouped and representative models are selected from each group. History matching is performed on the representative models instead of using all models to expedite the process of uncertainty quantification. Third, GTTI is implemented as an optimization algorithm in RML to take advantage of strengths of streamline based inversion such as computational efficiency and quasilinearity.

This paper is comprised of five chapters. Chapter 1 describes the research trends and applications of streamline simulation, streamline based history matching, and distance based approach. Chapter 2 explains theoretical backgrounds in streamline simulation, generalized travel time inversion, distance based method, and RML. In chapter 3, the methodology is proposed to quantify uncertainties with streamline based inversions and distance based method. Chapter 4 presents results of uncertainty assessment and comparison with previous methods. Finally, chapter 5 summarizes and concludes the study.

2. Theoretical backgrounds

2.1 Streamline simulation

A streamline simulator approximates 3-D flow calculations by a sum of 1-D calculations on streamlines (Datta-Gupta, 2000). Streamline simulator has several advantages over other conventional simulators. One of the advantages is that streamline method provides rapid computations. Datta-Gupta (2000) summarized four principle reasons why streamline simulation gives high computational efficiency. First, streamlines do not have to be updated frequently. Theoretically, streamlines change over time since the pressure distribution throughout the reservoir changes. However, the assumption that streamlines are fixed falls well within the boundary of uncertainty of the input data without changing reservoir conditions, such as infill drilling, shut-in of wells or changing flow rates (Thiele, 2005).

Second, the transport equations along streamlines are solved on 1-D TOF coordinate and often can be solved analytically. Third, since the flow equation is transformed to 1-D TOF coordinate, the grid stability condition can be alleviated, which means that the larger timesteps can be used (Batychy et al., 1997; Blunt et al., 1996, Jang and Choe, 2001; Cheng, 2005). Fourth, the rapidity of streamline simulation is prominent when the reservoir is highly heterogeneous. When the reservoir is strongly governed by heterogeneity, the computational time increases linearly with the number of gridblocks. Another advantage of streamline simulation is improved accuracy. Since the

heterogeneity is decoupled from flow calculations, it results in the reduction of numerical dispersion and grid-orientation effect (Datta-Gupta, 2000).

Streamline simulation for this study consists of the following steps. First, streamlines are traced using Pollock's algorithm. Second, 1-D transport equations in terms of TOF are solved. Third, saturations on the streamline are mapped back to the 2-D Cartesian coordinate. The following items summarize backgrounds of performing streamline simulation

1) Tracing streamlines and calculating TOF

The key principle of streamline simulation is that the partial differential equations of reservoir flows are solved on 1-D TOF coordinate instead of the Cartesian coordinate by transformation of coordinate. TOF is defined as Eq. (1).

$$\tau = \int \frac{d\xi}{v} = \int \frac{\phi}{u} d\xi \tag{1}$$

where, τ, v, ϕ, u, ξ mean TOF, interstitial velocity, porosity, Darcy velocity, spatial coordinate along the streamline, respectively. TOF physical means the transit time of a neutral tracer from an injector to a certain location. By representing physical distance by TOF, it incorporates permeability, porosity, and total mobility effects along the streamline (Batycky et al., 1997). Eq. (1) can be represented as the parametric form in Eq. (2).

$$\frac{d\tau}{\phi} = \frac{dx}{u_x} = \frac{dy}{u_y} \tag{2}$$

Pollock (1988) suggested the semi-analytical particle tracking method using velocities obtained from a finite difference model. Darcy velocities are obtained after pressure distributions are computed with IMPES (IMplicit Pressure Explicit Saturation). By assuming that the velocity within a cell varies linearly, the exit coordinate and TOFs of a particle within a cell are computed analytically. Here, the principles and procedures for applying Pollock's algorithm are described (Datta-Gupta and King, 2007).

As shown in Figure 1, a particle is initially located at (x_0, y_0). Since the linear slope of the velocity is assumed, it can be written as Eq. (3) and the velocity at an arbitrary location can be computed by Eq. (4).

$$c_x = \frac{u_{x2} - u_{x1}}{\Delta x}$$
$$c_y = \frac{u_{y2} - u_{y1}}{\Delta y} \tag{3}$$

$$u_x = u_{x1} + c_x (x - x_1)$$
$$u_y = u_{y1} + c_y (y - y_1) \tag{4}$$

where, c_x and c_y are the slope of Darcy velocity in x and y direction, respectively.

TOFs to reach each face are obtained by analytic integration from Eq. (2) and (4). In 2-D, we obtain four TOFs by Eq. (5).

$$\frac{\Delta \tau_{xi}}{\phi} = \int_{x_0}^{x_i} \frac{dx}{u_x} = \int_{x_0}^{x_i} \frac{dx}{u_{x0} + c_x (x - x_0)} = \frac{1}{c_x} \ln\left(\frac{u_{xi}}{u_{x0}}\right) \tag{5}$$

$$\frac{\Delta\tau_{yi}}{\phi}=\int_{y_0}^{y_i}\frac{dy}{u_y}=\int_{y_0}^{y_i}\frac{dy}{u_{y0}+c_y(y-y_0)}=\frac{1}{c_y}\ln\left(\frac{u_{yi}}{u_{y0}}\right)$$

where, subscript $i=1,2$ indicates each face.

TOF within a cell is the minimum positive TOF calculated by Eq. (5), which can be expressed as Eq. (6).

$$\Delta\tau=\text{Min Positive}\,(\Delta\tau_{x1},\Delta\tau_{x2},\Delta\tau_{y1},\Delta\tau_{y2}) \tag{6}$$

The exit coordinates are computed in Eq. (7) by substituting the TOFs from Eq. (6). The exit coordinates from Eq. (7) are now the entry coordinate for the next cell. This is repeated until the particle reaches the producing well.

$$\begin{aligned}x_e&=x_0+u_{x0}\left(\frac{e^{c_x\Delta\tau/\phi}-1}{c_x}\right)\\ y_e&=y_0+u_{y0}\left(\frac{e^{c_y\Delta\tau/\phi}-1}{c_y}\right)\end{aligned} \tag{7}$$

2) Transformation of the flow equation to the TOF coordinate

Black oil equations transformed to the TOF coordinate can be written as Eq. (8) for water and Eq. (9) for gas (Adedayo, 2008).

$$\phi\frac{\partial}{\partial t}\left(\frac{S_w}{B_w}\right)+\phi\frac{\partial}{\partial\tau}\left(\frac{f_w}{B_w}\right)=-c\frac{f_w}{B_w} \tag{8}$$

$$\phi \frac{\partial}{\partial t}\left(\frac{S_g}{B_g} + \frac{S_o R_s}{B_o} \right) + \phi \frac{\partial}{\partial \tau}\left(\frac{f_g}{B_g} + \frac{R_s f_o}{B_o} \right) = -c\left(\frac{f_g}{B_g} + \frac{R_s f_o}{B_o} \right) \quad (9)$$

where, S, B, f are the saturation, formation volume factor, and fractional flow, respectively. The subscripts w, o, g mean water, oil, and gas phases. R_s is the solution gas-oil ratio and c represents the divergence of flux, which can be written as Eq. (10) (Datta-Gupta and King, 2007).

$$\nabla \bullet u = \sum_{j=1}^{3} c_j = c_x + c_y + c_z \quad (10)$$

Assuming incompressible flow, c is zero at Eq. (10). This simplifies Eq. (8) into Eq. (11).

$$\frac{\partial S_w}{\partial t} + \frac{\partial f_w}{\partial \tau} = 0 \quad (11)$$

Saturations on the streamline are obtained by solving Eq. (11). There are several methods to solve Eq. (11) and one of them is to use the analogy from the analytic solution of Buckley-Leverette equation. This characteristic contributes to the computational efficiency and stability of streamline simulation. In addition, it is clearly seen in Eq. (11) that the reservoir heterogeneity is decoupled from the Cartesian grid.

3) Mapping saturations from streamlines to Cartesian grid

Saturations computed on TOF coordinate are mapped to the original Cartesian coordinate. For incompressible fluid, saturation is computed as a weighted average as shown in Eq. (12). If compressibility is considered, it can be mapped by introducing effective density ρ, as in Eq. (13) (Cheng et al., 2006).

$$\overline{S}_{gridblock} = \frac{\sum_i q_i \Delta\tau_i S_i}{\sum_i q_i \Delta\tau_i} \tag{12}$$

$$\overline{S}_{gridblock} = \frac{\sum_i q_i \Delta\tau_i S_i / \rho_i}{\sum_i q_i \Delta\tau_i / \rho_i} \tag{13}$$

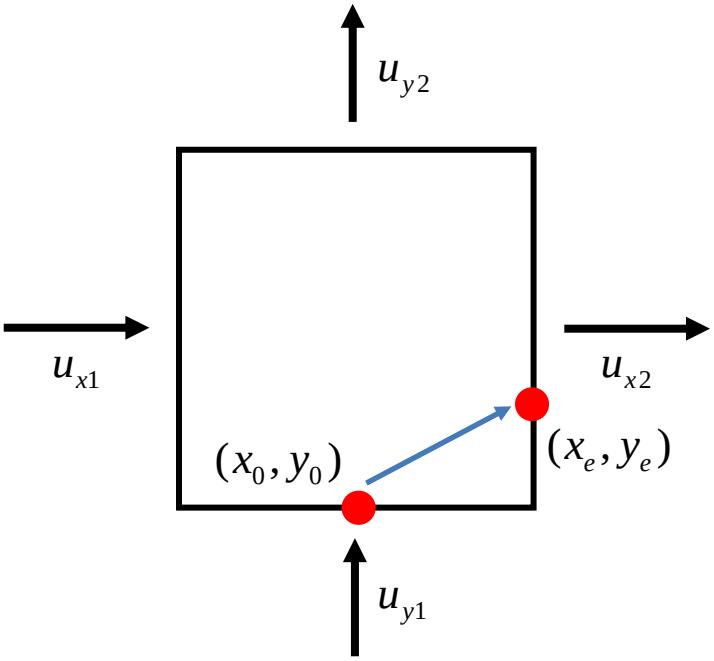


Figure 1 Streamline tracing in a gridblock

2.2 Generalized travel time inversion

2.2.1 Generalized travel time

In conventional history matching, data misfit is defined as the difference between observed data and model responses. The square sum of the differences is a part of the objective function, and the objective function is minimized via various optimization techniques. This is referred as amplitude inversion. In contrast, in travel time inversion and GTTI, the data misfit is defined as time. In travel time inversion, the data misfit is set to be the difference of breakthrough time or time when the maximum concentration occurs.

In GTTI, data misfit is defined as GTT. GTT is an optimal time shift that minimizes the square sum of the amplitude misfit between observed data and model responses. For well j, it is either $\Delta\tilde{t}_j$ that minimizes the objective function J in Eq. (14) or maximizes the coefficient of determination R expressed in Eq. (15). Then, the overall data misfit can be computed with Eq. (16) (He et al., 2002; Cheng et al., 2005).

$$J = \sum_{i=1}^{N_{dj}} [y_j^{obs}(t_i + \Delta t_j) - y_j^{cal}(t_i)]^2 \tag{14}$$

$$R^2(\Delta t_j) = 1 - \frac{\sum [y_j^{obs}(t_i + \Delta t_j) - y_j^{cal}(t_i)]^2}{\sum [y_j^{obs}(t_i) - \bar{y}_j^{obs}]^2} \tag{15}$$

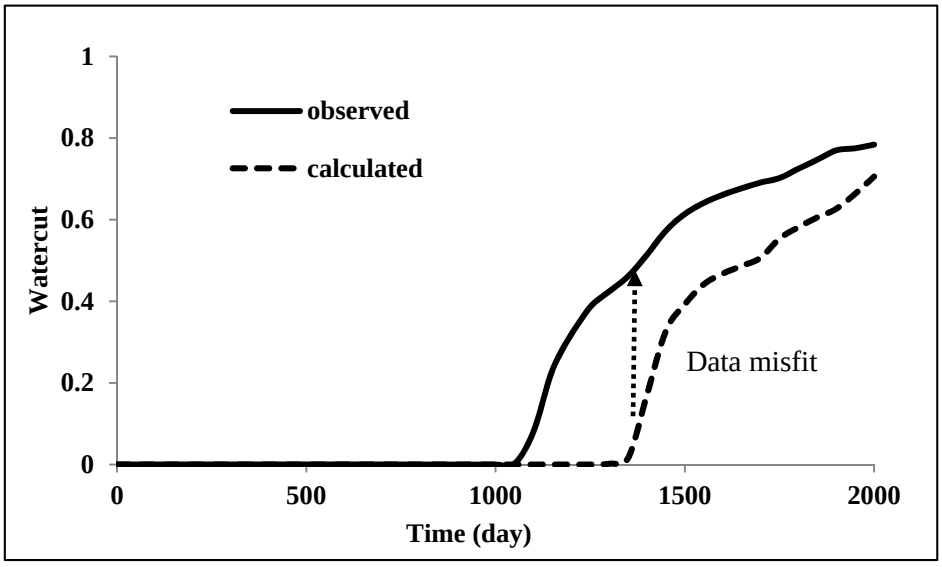
$$E = \sum_{j=1}^{N_w} (\Delta t_j)^2 \tag{16}$$

where, t, Δt, y, N_w, N_{dj} mean the time, shifting time, production response, the number of production wells, and the number of data points of well j, respectively. Superscript obs, cal mean observed and calculated values. The bar means an averaged value.

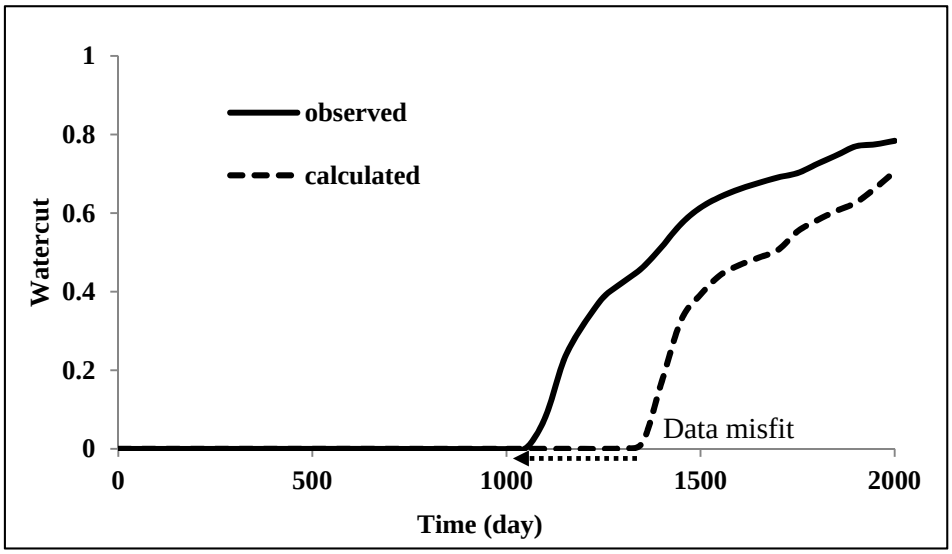
Figure 2 compares the amplitude inversion and the travel time inversion. In Figure 2, the solid line is the observed watercut and the dashed line is the watercut calculated from the model. As seen in Figure 2a, the data misfit is the difference of the production responses in amplitude inversion. Figure 2b shows that the data misfit is the differences in breakthrough times.

Figure 3 shows how GTT is determined. Using Eq. (15), the coefficient of determination is calculated for each shifting time. In this example, the value of shifting time -271 day maximizes the coefficient of determination. It indicates that the amplitude differences between the observed data and model responses are minimized if we shift the calculated responses by 271 days to the left, as illustrated in Figure 3b.

The advantage of quantifying the data misfit as time is quasilinearity. Inversion problems in reservoir simulations are typically highly nonlinear, so results of inversion are sensitive to the choice of a prior model. The quasilinearity of travel time inversion makes the inversion less sensitive to the choice of an initial model, and the objective function converges rapidly. GTTI can also take amplitude differences into account while maintaining the strengths of travel time inversion.

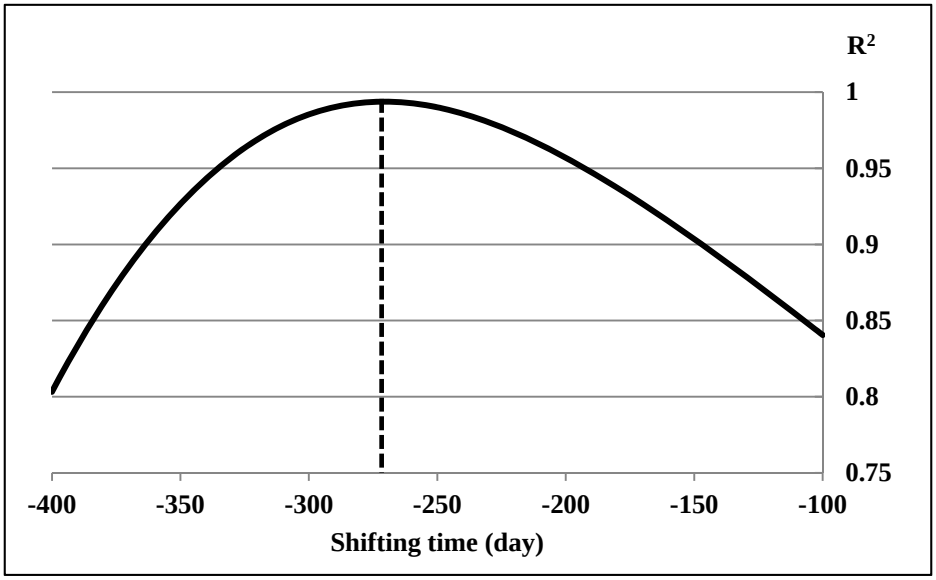


(a) Data misfit in amplitude inversion

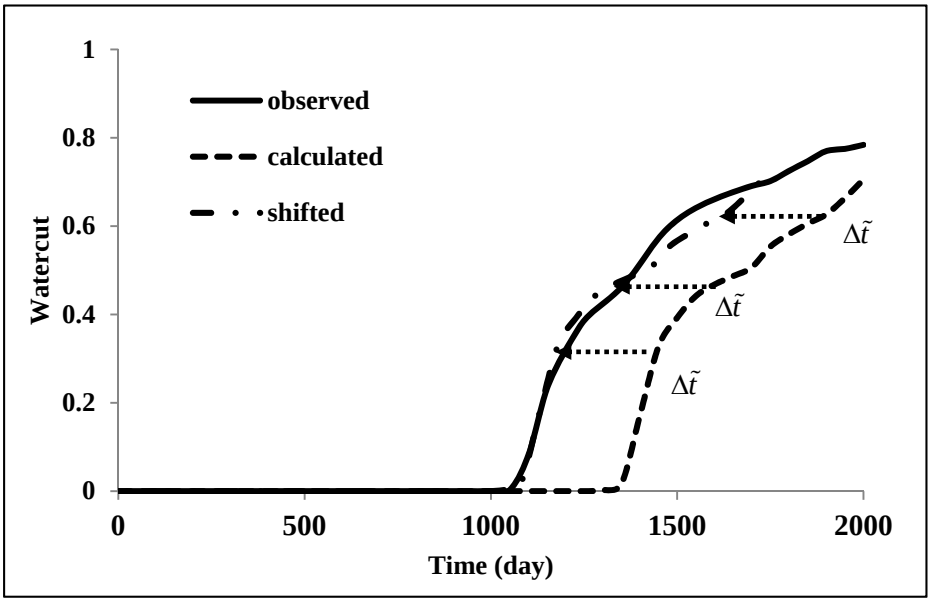


(b) Data misfit in travel time inversion

Figure 2 Quantifying data misfit in amplitude inversion and travel time inversion



(a) Estimating GTT by coefficient of determination



(b) Data misfit in GTTI

Figure 3 Estimation of GTT and data misfit in GTTI

2.2.2 Sensitivity calculation

He et al. (2001) derived travel time sensitivity as follows. Assuming incompressible flow, the transport equation can be written as Eq. (11), and can be rearranged to Eq. (17).

$$\frac{\partial S_w}{\partial t} = -\frac{\partial f_w}{\partial S_w}\frac{\partial S_w}{\partial \tau} \tag{17}$$

Assuming that the streamline does not shift, the change of water saturation at the streamline outlet node can be expressed in the differential form in Eq. (18). Here, m is a vector of model parameter.

$$\delta S_w = \frac{\partial S_w}{\partial t}\delta t + \frac{\partial S_w}{\partial \tau}\left[\frac{\partial \tau}{\partial m}\right]^T \delta m \tag{18}$$

Letting δS_w be zero in Eq. (18) to express fixed propagation of saturation and substituting the equation to Eq. (17), travel time sensitivity is obtained as Eq. (19).

$$\frac{\partial t}{\partial m} = \frac{\dfrac{\partial \tau}{\partial m}}{\dfrac{\partial f_w}{\partial S_w}} \tag{19}$$

The saturation speed in Eq. (19) can be computed at the outlet node of the streamline, and TOF sensitivity with respect to the model parameter can be computed analytically in terms of integrals along the streamline as follows.

The slowness, which is reciprocal of interstitial speed, is defined as Eq. (20)

$$s(x) = \frac{1}{|v(x)|} = \frac{\mu\phi(x)}{k(x)|\nabla P|} \tag{20}$$

where, μ, ϕ, P are viscosity, porosity and pressure, respectively.

Then, TOF can be defined as the integral of the slowness as Eq. (21). Using this definition, TOF sensitivity with respect to permeability and porosity is computed with Eq. (22).

$$\tau = \int_{\psi} s(x)dr \tag{21}$$

$$\begin{aligned} \frac{\partial \tau}{\partial k(x)} &= \int_{\psi} \frac{\partial s(x)}{\partial k(x)} dx = -\int_{\psi} \frac{s(x)}{k(x)} dx \\ \frac{\partial \tau}{\partial \phi(x)} &= \int_{\psi} \frac{\partial s(x)}{\partial \phi(x)} dx = \int_{\psi} \frac{s(x)}{\phi(x)} dx \end{aligned} \tag{22}$$

Assuming that each reservoir parameter is constant within a cell, the TOF sensitivity with respect to permeability and log permeability at a certain cell can be calculated as Eq. (23) from Eq. (22).

$$\begin{aligned} \frac{\partial \tau}{\partial k} &= -\frac{\Delta \tau}{k} \\ \frac{\partial \tau}{\partial \ln k} &= -\Delta \tau \end{aligned} \tag{23}$$

GTT sensitivity can be computed with the method suggested by He et al. (2001). Let's assume that there is a small perturbation on model parameters and the entire data points have the identical time shift $\Delta\tilde{t}_j$ for well j. Then, the change in TOF can be expressed with respect to observation times $i = 1, \cdots, N_{dj}$ as expressed in Eq. (24)

$$\delta t_j = \delta t_i = \left[\frac{\partial t_{N_d, j}}{\partial m} \right]^T \delta m \tag{24}$$

Summing Eq. (24) over $i = 1, \cdots, N_{dj}$ gives Eq. (25).

$$\frac{\partial t_j}{\partial m} = \frac{\sum_{i=1}^{N_{dj}} \left(\frac{\partial t_{i,j}}{\partial m} \right)}{N_{dj}} \tag{25}$$

From the definition of GTT, GTT sensitivities are obtained as Eq. (26).

$$\frac{\partial \Delta\tilde{t}_j}{\partial m} = -\frac{\partial t_j}{\partial m} = -\frac{\sum_{i=1}^{N_{dj}} \left(\frac{\partial t_{i,j}}{\partial m} \right)}{N_{dj}} \tag{26}$$

Eq. (26) shows that GTT sensitivities can be computed using the average of travel time sensitivities.

2.2.3 Data Integration

There are two approaches defining an objective function in streamline based automatic history matching – deterministic and Bayesian approaches. It is proved that both of the methods have equal success in history matching (Cheng et al., 2005).

In deterministic approach, an objective function is defined as Eq. (27).

$$O(m) = \left\| \Delta\tilde{t} - G\delta m \right\| + \beta_1 \left\| \delta m \right\| + \beta_2 \left\| L\delta m \right\| \quad (27)$$

where, β_1, β_2 are the weightings, and L is the discrete model Laplacian, a finite difference approximation to the second spatial derivative of the deviation (Datta-Gupta and King, 2007).

The first term in Eq. (27) indicates the data misfit between the observed data and model responses. There are two penalized terms in Eq. (27). The second term is a norm constraint, which penalizes the deviation of the model to preserve the geological realism. The third term, roughness penalty, recognizes the fact that the history matching is best suited to resolve large-scale structures rather than small-scale property variations (Cheng, 2005). The drawbacks of the deterministic approach is that the inversion result is sensitive to the choice of weightings β_1, β_2 and it is not possible to assess the uncertainty.

In Bayesian approach, an objective function is defined as Eq. (28). Here, C_M, C_d, and m_{pr} are prior model covariance matrix, data error covariance matrix, and prior model parameter, respectively. C_M can be constructed with geostatistical parameters

such as variogram. The advantage of Bayesian approach is that it is possible to assess the uncertainty with multiple realizations. So in this paper, Bayesian approach is adopted.

$$O(m) = \frac{1}{2}[(m - m_{pr})^T C_M^{-1}(m - m_{pr}) + (\Delta\tilde{t})^T C_D^{-1}(\Delta\tilde{t})] \quad (28)$$

The objective functions from Eq. (27) or Eq. (28) can be minimized through several minimization algorithms, such as Gauss-Newton iteration, Levenberg-Marquardt method, or LSQR (sparse linear equations and sparse least squares) algorithm. In this study, Gauss-Newton iteration is used. In Gauss-Newton iteration, the following recursive equations in Eq. (29) are used to update model parameters (Oliver et al., 2008).

$$\begin{aligned} &\delta m^{l+1} = -(m^l - m_{pr}) \\ &-C_M G_l^T (C_D + G_l C_M G_l^T)^{-1}(g(m^l) - d_{obs} - G_l(m^l - m_{pr})) \\ &m^{l+1} = m^l + \delta m^{l+1} \end{aligned} \quad (29)$$

2.2.4 Procedures of generalized travel time inversion

Figure 4 is the flow chart to perform GTTI (Datta-Gupta and King, 2007). Starting from a prior geological model, production responses are obtained from a forward simulator. The simulator can be either a streamline simulator or a finite difference simulator. This is another advantage of streamline based automatic history matching - the algorithm is compatible with previously developed simulators without modifications on them.

After dynamic responses are obtained, data misfit is calculated as discussed in Section 2.2.1. After generating streamlines, streamline based sensitivities are computed with TOFs and other related variables. If a streamline simulator is utilized as a forward simulator, they have been already secured in the previous step. If other simulators are used, streamlines are generated with pressure and flux obtained using the procedures discussed in Section 2.1.

Sensitivities are computed and the sensitivity matrix is constructed to use in the optimization. The objective function defined in either Eq. (27) or Eq. (28) is minimized with various optimization algorithms. The model parameter is now updated and serves as a prior for the next iteration. This process is repeated until the objective function meets a given convergence criterion.

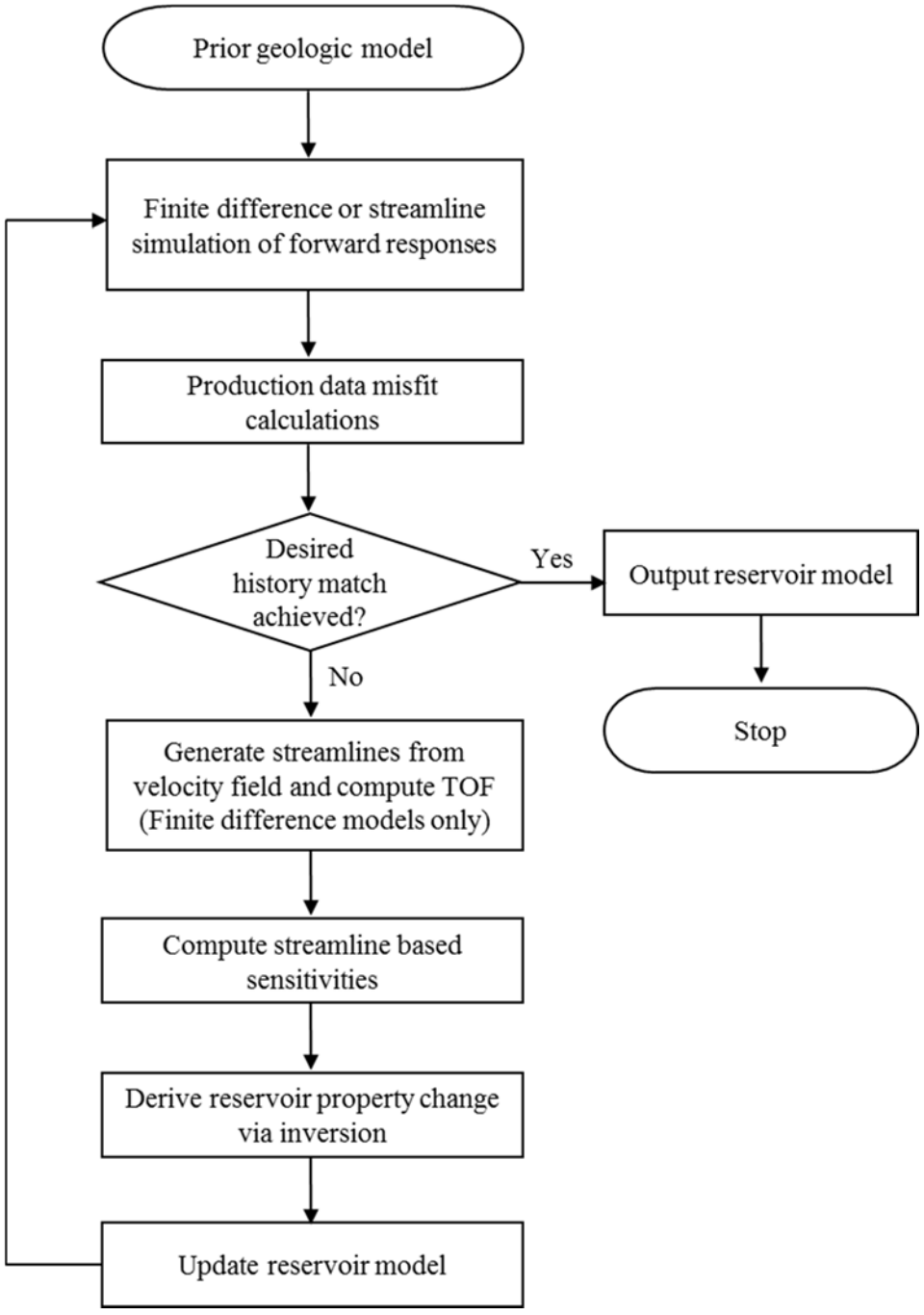


Figure 4 Steps for GTTI (Datta-Gupta and King, 2007)

2.3 Distance based method

2.3.1 Concept of distance

Uncertainty quantification of reservoir performances is carried out via generating a lot of alternative reservoir models and running flow simulations and inversions, which require high computational cost. Taking distance based approach, we can quantify uncertainties with a few selected models according to a ‘distance’ between the models.

A distance is a quantitative measure of differences between each model. A distance can be calculated in any manner, as long as it is correlated to the flow response of interest (Schedit and Caers, 2009). This is illustrated in Figure 5. In Figure 5, we see that models with close distance have similar flow responses. This can be written in Eq. (30) if we use the Euclidean distance. If there are N multiple reservoir models, we can construct a $N \times N$ distance matrix.

$$d_g(g(m_i), g(m_j)) = \sqrt{(g_i - g_j)^T (g_i - g_j)} \tag{30}$$

where, d , g are observed data and a forward simulator, respectively.

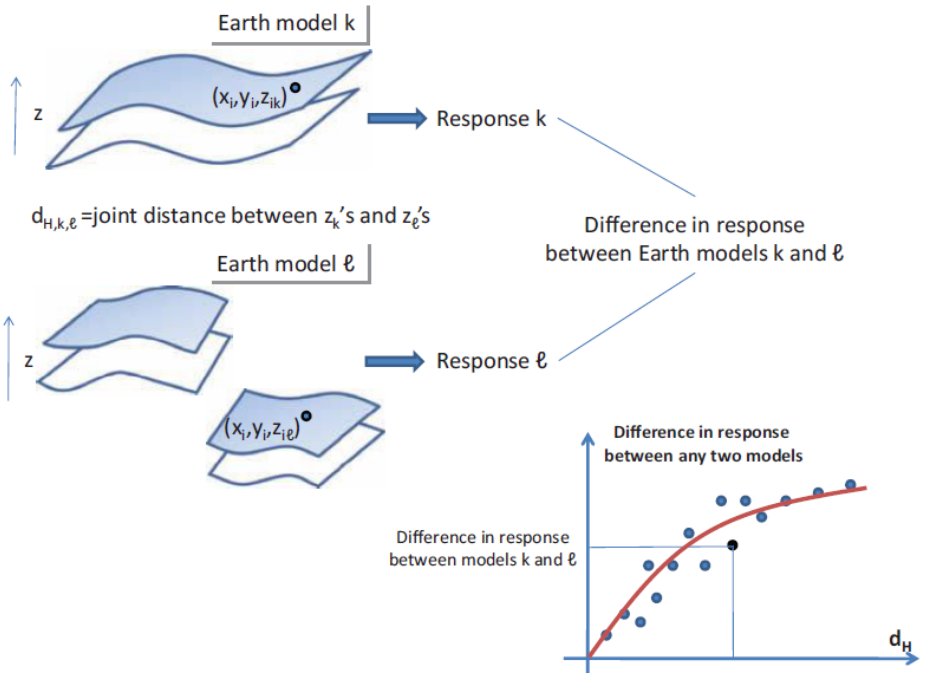


Figure 5 Correlation between the difference in the flow response and distance

(Caers, 2011)

2.3.2 k-means clustering

After models are scattered according to their mutual distances, they need to be grouped to form each cluster. Within a cluster, a representative model is chosen. There are several methods in clustering, and k-means clustering is widely used. The goal of k-means clustering is to cluster n objects into k classes. The procedures for k-means clustering is as follows (Caers, 2011):

1) Set n cluster centers randomly in the space.
2) Calculate the distance between centers and each model.
3) Assign models to the closest centers.
4) Calculate new means to obtain new centers based on the assigned models.
5) Go to step 2 until there is no change in cluster centers.

After models are grouped into k clusters, each model that is closest to the mean of the cluster is selected as a representative model.

2.4 Randomized maximum likelihood

Randomized maximum likelihood method is one of the techniques to quantify uncertainties by generating realizations from multi-normal distributions or for independent stochastic variables (Oliver et al., 2008). It is suggested by Kitanidis (1995) and Oliver et al. (1996). It consists of three steps, and can be summarized as follows (Oliver et al., 2008; Cheng, 2005).

1) Generate an unconditional realization of model parameters m_u from $N(m_{pr}, C_M)$. This can be done using a Cholesky decomposition of C_M for small problems, and sequential Gaussian simulation for larger problems (Gao et al., 2005).
2) Generate an unconditional realization of dynamic data d_u from $N(m_{pr}, C_M)$. This is done by adding noise $\varepsilon \sim N(0, C_D)$ to the model response, that is, $d = g(m) + \varepsilon$
3) Compute the conditioned model parameter m_c that minimizes the objective function in Eq. (31).

$$O(m) = \frac{1}{2}[(m - m_u)^T C_M^{-1}(m - m_u) + (d_u - g(m))^T C_D^{-1}(d_u - g(m))] \qquad (31)$$

Instead of using a single prior model and observed data, the minimization is carried out with respect to unconditional realization of the model and data. As a result, the

conditional realizations from the minimization are unbiased and offer multiple results to quantify uncertainties (Jung, 2008).

The objective function in Eq. (31) should satisfy the following criteria for convergence expressed in Eq. (32).

$$N_d - 5\sqrt{2N_d} \le O(m_c) \le N_d + 5\sqrt{2N_d} \tag{32}$$

3. Quantifying uncertainty with GTTI, RML, and distance based method

Figure 6 summarizes Cheng (2003)'s method to quantify uncertainty and Figure 7 illustrates the procedures for the proposed method. In both of the methods, one hundred realizations are generated via sequential Gaussian simulation (SGS). Cheng (2003) computed conditional model parameters by substituting 100 prior models to Eq. (31) and minimizing the objective function with GTTI. In the proposed method, instead of matching all 100 models, representative models are conditioned after they are extracted via distance based method.

The first step in distance based approach is to define a distance. The distance of each model is defined as the norm of differences of GTT vectors in this research. The motivation comes from the fact that the observed data are time instead of production responses. This also mathematically corresponds to how the data misfit is defined in GTTI as seen in Eq. (28).

Figure 8 illustrates the idea of defining the distance as GTT. In Figure 8, the observed watercut is plotted in a solid line, and two watercut responses from two models which are named model A, B are plotted with a squared dotted line and a dotted line, respectively. GTT vectors of the model A, B are $\Delta\tilde{t}_A$ and $\Delta\tilde{t}_B$. Then, the distance between model A and model B can be defined as a norm between $\Delta\tilde{t}_A$ and $\Delta\tilde{t}_B$ because it is previously assumed that each point on the watercut curve shifts by the same amount of time. As a result, the distance can be formulated as Eq. (33).

$$d_{ij} = \left\| \Delta\tilde{t}_i - \Delta\tilde{t}_j \right\| \tag{33}$$

The number of elements in GTT vector is the same as the number of producing wells. Therefore, if each GTT vector is recognized to be located at N_w- space, the distance defined in Eq. (33) corresponds to the Euclidean distance in Eq. (30). For example, in this research, the field has eight producing wells, so we can regard that the generated models are distributed in 8-D metric space according to their GTT. This is illustrated in Figure 8.

After computing all distances, k-means clustering is applied to make clusters and select the representative models. In this research, ten clusters are used and representative models are selected from each cluster. RML is applied to the selected models to quantify uncertainties. Then the result of the proposed method is compared to that of Cheng (2003)'s method.

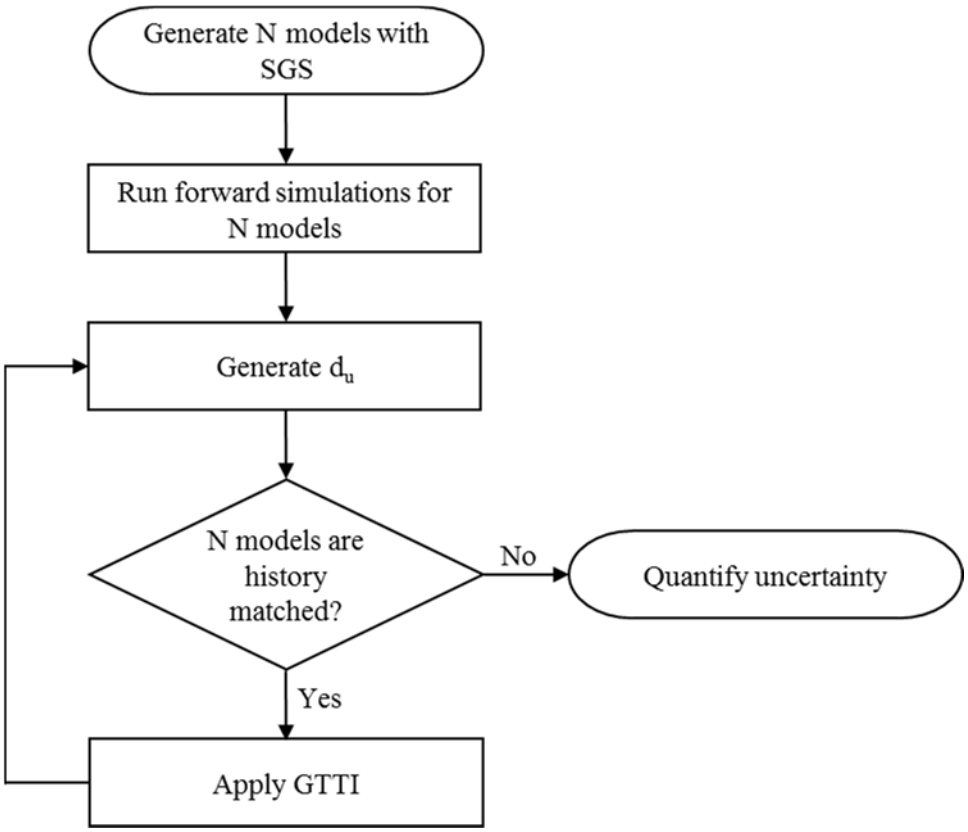


Figure 6 Procedures for applying RML to all initial models

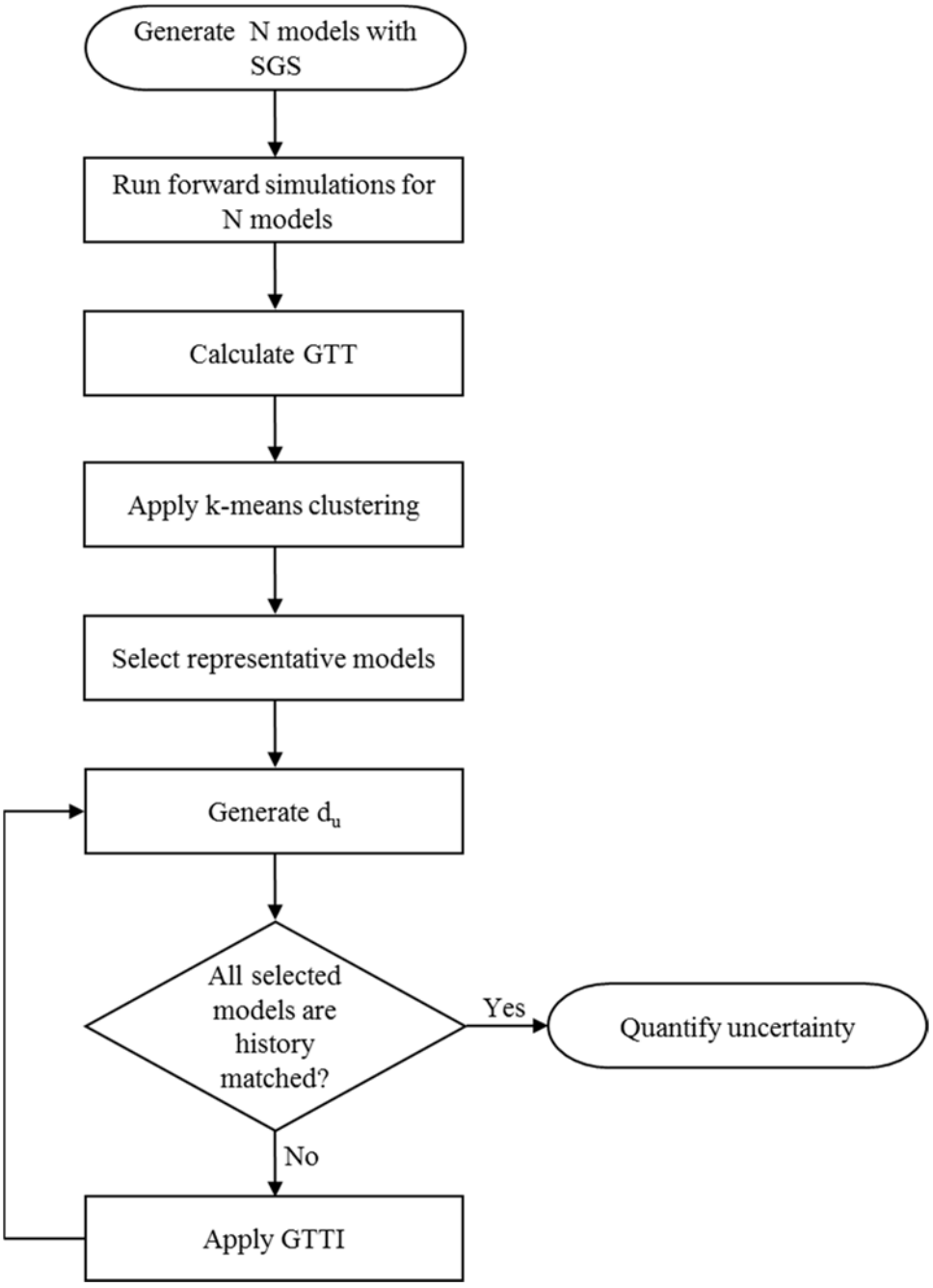


Figure 7 Procedures for the proposed method

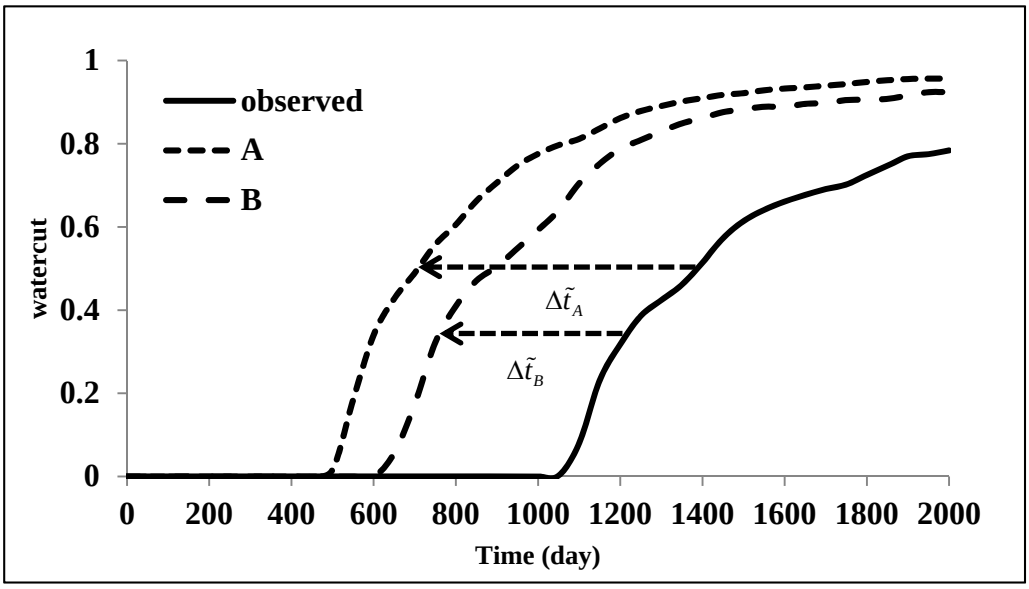


Figure 8 Estimating distance as a norm of GTT difference

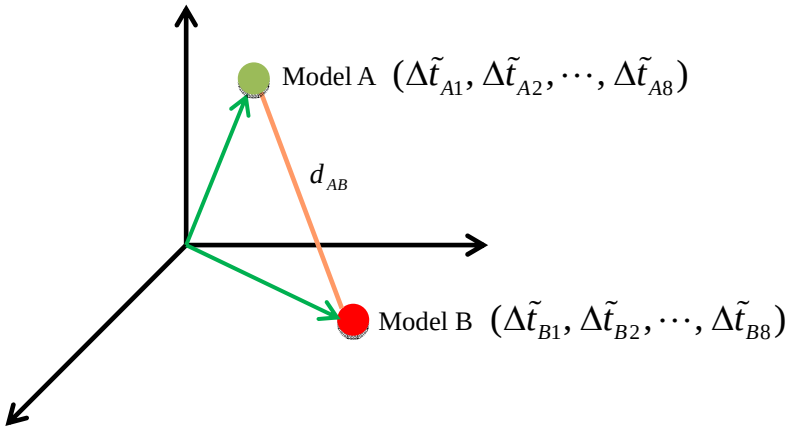


Figure 9 GTT vectors in multidimensional space to define the distance

4. Results

4.1 Reference field

The reference field is a 2-D square synthetic field of which the dimension is 1,050 ft × 1050 ft × 20 ft, as shown in Figure 10. It consists of 21 × 21 × 1 cells, total 441 cells. Table 1 offers grid properties. As seen in Figure 10, highly permeable streak stretches from the upper-right to the lower-bottom. On the other hand, there is low permeability region on the left part of the reservoir, which also can be identified in Figure 11, where streamlines are generated. It can be observed that the density of streamlines is higher in high permeability region. Since fluid behaviors are sensitive to the regions of comparatively high and low permeability regions, it is significant to spot those regions in reservoir characterization (Choe, 2007; Jung, 2008; Yeo, 2010). It is assumed that permeability is isotropic in x and y directions. There are two kinds of reservoir fluids (oil and water) and Table 2 gives fluid properties. Figure 12 shows the relative permeability of each fluid with respect to water saturation.

Geostatistical information of the reference field is summarized in Table 3. It is assumed that the permeability follows log normal distribution. The average and standard deviation are 3.83 and 0.40, respectively. It is also assumed that the permeabilities of the cells where wells are located are known. SGS is applied to generate log permeability of rest of cells. mGstat, a geostatistical Matlab toolbox, is utilized to generate data with SGeMS (Stanford Geostatistical Modeling Software) and process them on Matlab

platform. A spherical model is adopted for variogram, and it is calculated as Eq. (34). This is used in data generation with SGS, and construction of a prior covariance matrix.

$$\gamma(h) = C_0 Sph_a(h) = \begin{cases} C_0\left[1.5\left(\frac{h}{a}\right) - 0.5\left(\frac{h}{a}\right)^3\right], \ for\ h \le a \\ C_0, \ for\ h > a \end{cases} \tag{34}$$

where, γ, h, C_0, a are variogram, separation distance, sill, and range, respectively. Specific values are given in Table 3.

The reservoir is waterflooded with inverted nine-spot pattern. There is an injection well at the center of the square, four producing wells at the corners and another four wells at the middle of four sides. Table 4 summarizes well information. Each producing well has constant bottomhole pressure condition of 500 psia. In the injection well, 300 STB/day of water is injected. Fluids are produced for 2,000 days, and Frontsim, a commercial streamline simulator is utilized.

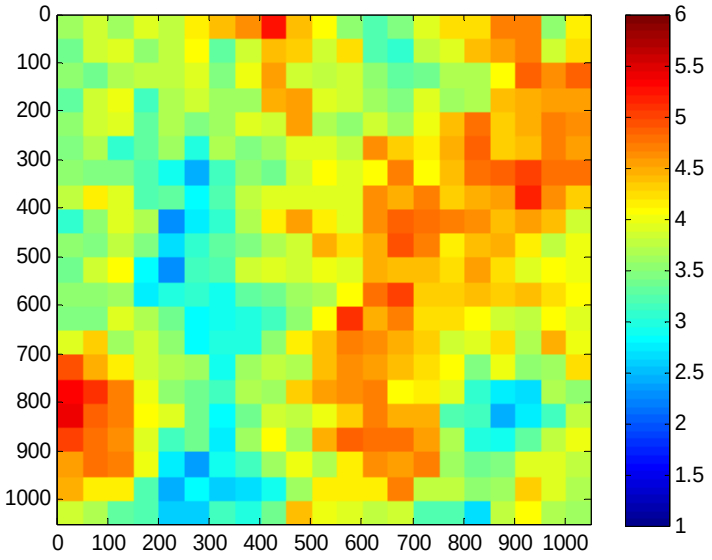


Figure 10 Log permeability distribution of the reference field

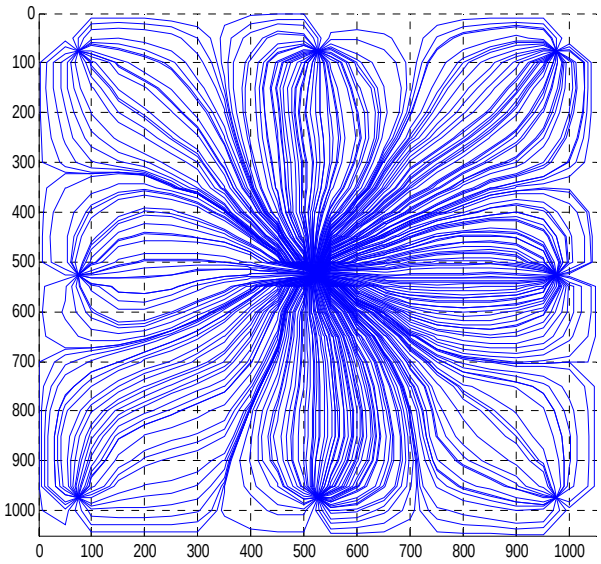


Figure 11 Generated streamlines

Table 1 Grid properties of the reference field

Parameters	**Value**
Number of gird cells	21 × 21 × 1
Width of the reservoir, ft	1,050
Length of the reservoir, ft	1,050
Thickness of the reservoir, ft	20
Width, ft	50
Length, ft	50
Thickness, ft	20
Porosity, fraction	0.15
Initial pressure, psia	2,000

Table 2 Fluid properties of the reference field

Parameters	**Value**
Connate water saturation, fraction	0.25
Residual oil saturation, fraction	0.2
End-point relative water permeability, fraction	0.3
End-point relative oil permeability, fraction	0.7
Formation volume factor of water	1
Water compressibility at 2,000 psia , /psi	5.00E-7
Water viscosity, cp	1
Oil viscosity, cp	3

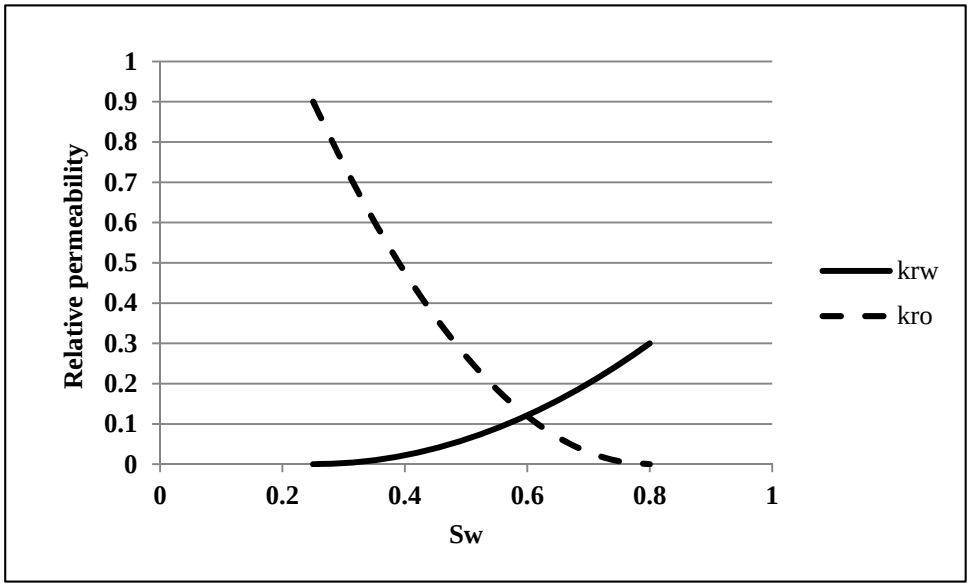


Figure 12 Relative permeability curves

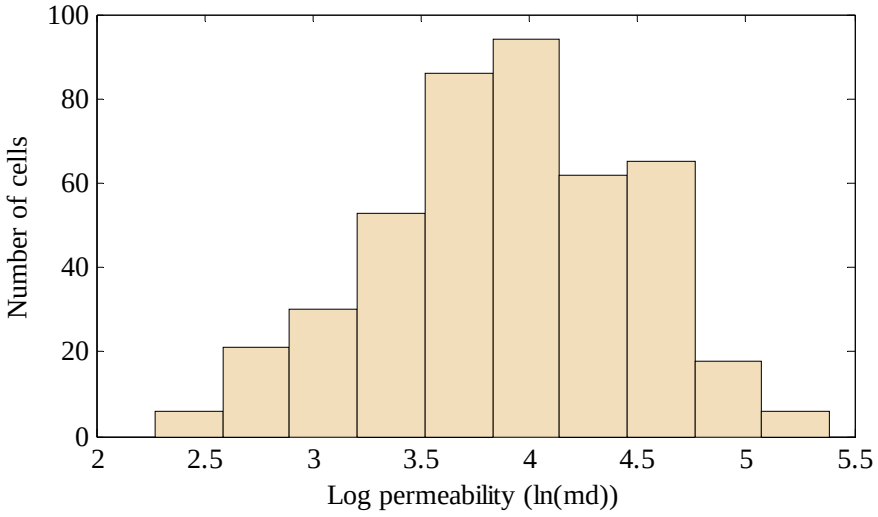


Figure 13 Histogram of log permeability

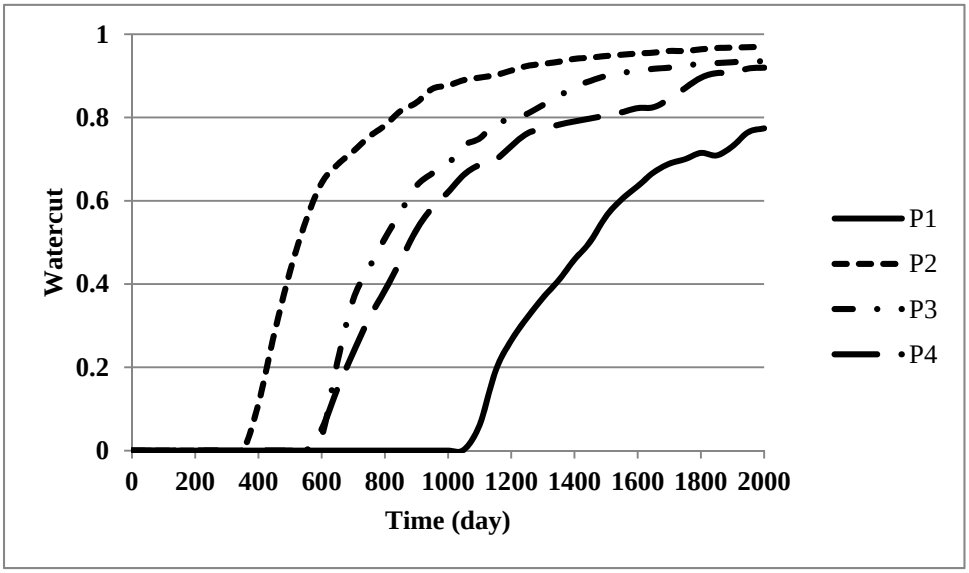
Table 3 Geostatistical data

Parameters	**Value**
Average log permeability, ln(md)	3.83
Standard deviation of log permeability, $ln(md)^2$	0.40
Varigogram model	Spherical
Sill	0.5
Nugget	0
Range, ft	500

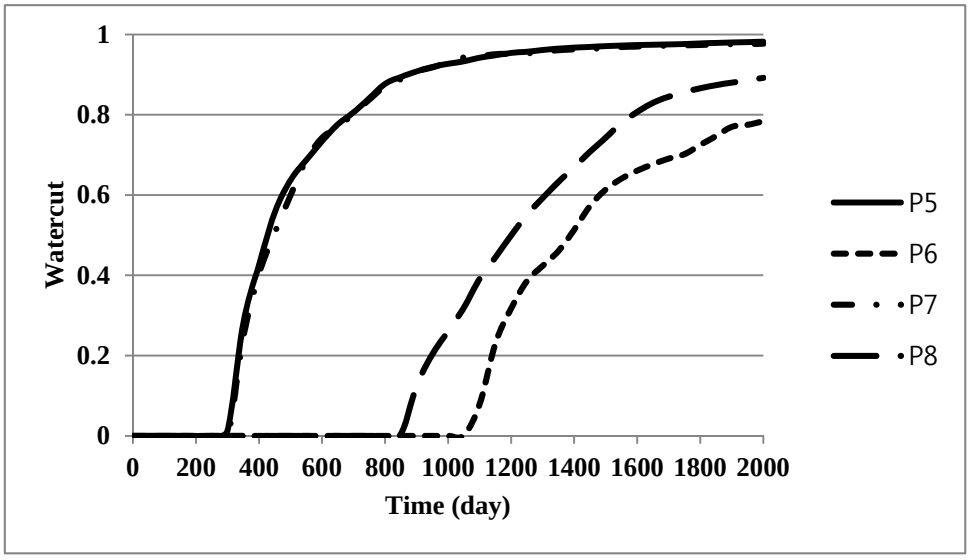
Table 4 Well data

Well name	**Well coordinate**	**Log permeability, ln(md)**	**Boundary condition**
P1	(2,2)	3.83	500 psia, constant bottomhole pressure
P2	(11,2)	3.85	
P3	(20,2)	3.87	
P4	(2,11)	3.84	
P5	(20,11)	4.09	
P6	(2,20)	4.20	
P7	(11,20)	4.15	
P8	(20,20)	3.90	
INJ	(11,11)	4.03	300 bbl/day, constant injection rate

Reservoir performances from a forward simulation are as follows. Figure 14 shows watercut from each producing well. It is expected that there will be early breakthrough at P2, P3, P5 and P7 wells due to the high permeability streak as seen from Figure 10. It is also anticipated that the last breakthrough occurs at P1, because the physical distance is the farthest and there is a low permeability region from the injection well. The result also shows that the breakthrough occurs at about 1,050 days, which is the most delayed. When breakthrough occurs, the production rate sharply decreases after breakthrough. Therefore, it is imperative to accurately match the time of breakthrough for the reliable reservoir characterization.



(a) Watercut from P1 to P4



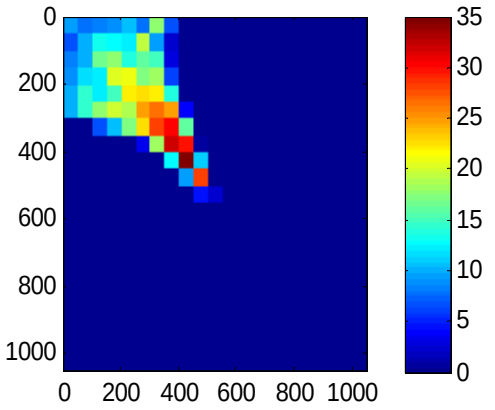
(b) Watercut from P5 to P8

Figure 14 Watercut from the reference field

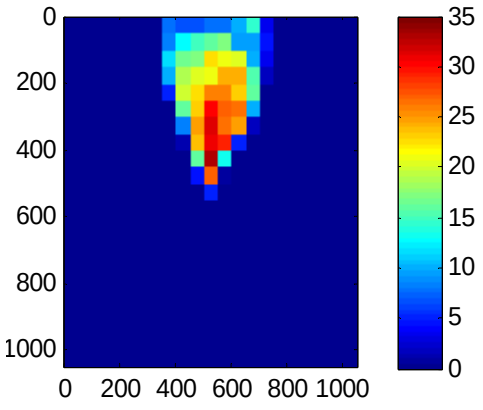
4.2 Sensitivity calculations

In streamline based automatic history matching, the sensitivity matrix G should be constructed at each iteration. Travel time sensitivity with respect to the model parameter can be estimated by Eq. (19). With travel time sensitivities acquired, GTT sensitivities are computed with Eq. (23).

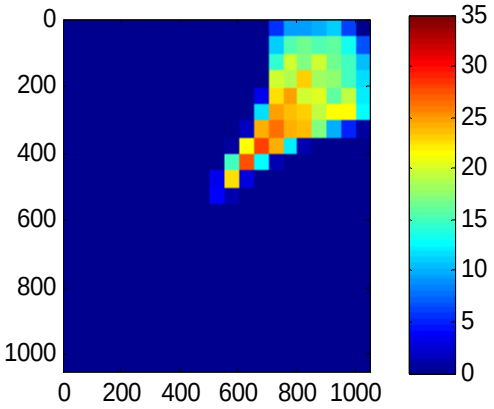
For the verification of the algorithm, travel time sensitivities at each producing well of a certain model are plotted in Figures 15 and 16. As expected, they exhibit ‘sensitivity band’ between the injection well and each producer, and zero values at other region. To be specific, Figure 15a displays travel time sensitivity map for P1. It has nonzero sensitivities on the region where streamlines connect the injection well and P1, and zero for other regions. This is a key characteristic of streamline based inversion – model parameters are updated along the path of the streamline. From the result of sensitivity map, the computed sensitivities are validated.



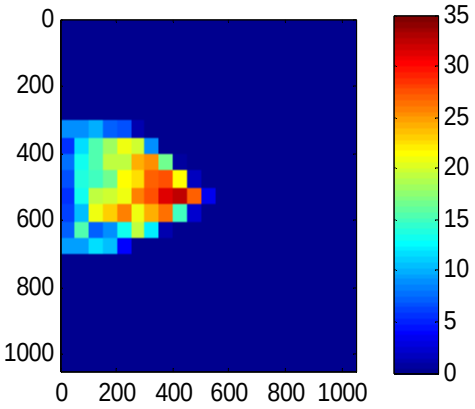
(a) Travel time sensitivity of P1



(b) Travel time sensitivity of P2

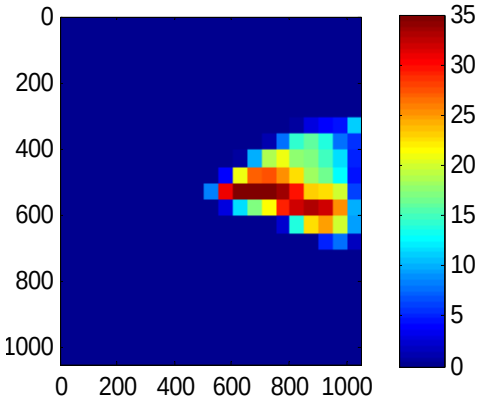


(c) Travel time sensitivity of P3

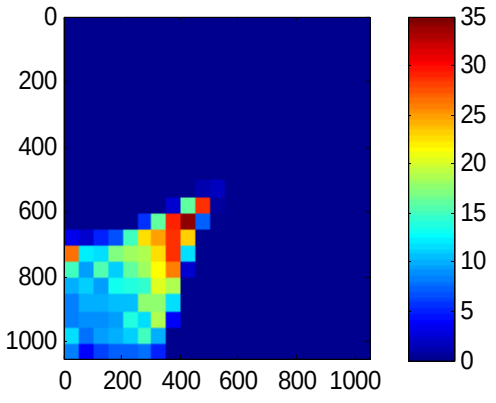


(d) Travel time sensitivity of P4

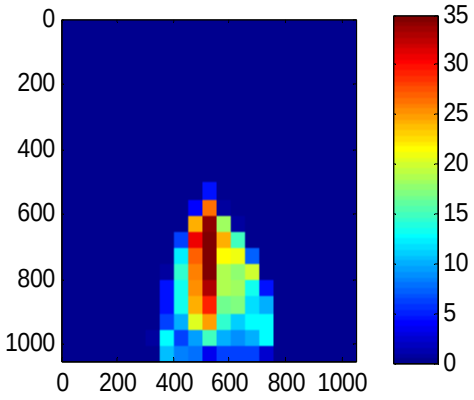
Figure 15 Travel time sensitivity of P1 to P4



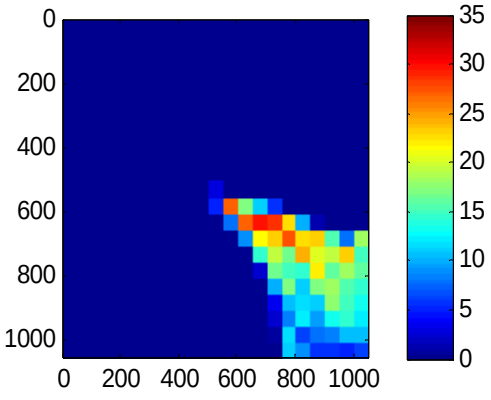
(a) Travel time sensitivity of P5



(b) Travel time sensitivity of P6



(c) Travel time sensitivity of P7



(d) Travel time sensitivity of P8

Figure 16 Travel time sensitivity of P5 to P8

4.3 Application of distance based method, RML and GTTI

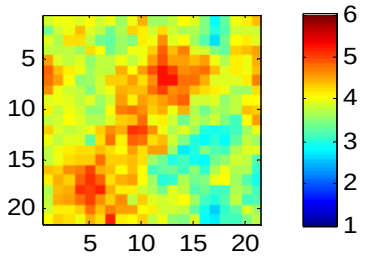
A hundred models are generated, and they follow geostatistical information from Table 4. Then, GTT vectors are calculated for all of them. If we regard GTT vectors as the vectors in multidimensional space, the distance is defined as the norm of the Euclidean distance as expressed in Eq. (31). By defining the distance this way, it is possible to capture well-specific characteristics. K-means clustering is applied to create clusters and select representative models with Matlab built-in algorithm. The results of initial and matched models are plotted in Figures 17 to 19. Here, the numbers 1, 2 mean the initial and the matched model.

It is observed that the initial model is corrected and the log permeability distribution that is similar to the reference field is found out via GTTI. As mentioned previously, the reference field is characterized as high permeability streak on the right, and low permeability region on the left. However, the 1st representative model has the high permeable region stretched from the lower left to the upper right, and lower permeability region on the east, which is far from the reference field. However, the model is corrected after inversion.

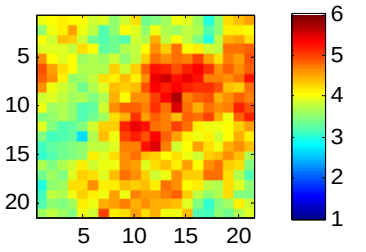
Similar analyses can be applied to other models except the 3rd representative model. The 5th representative model has high permeable streak from the upper left to the lower right and the 8th representative model has horizontal streak on the bottom. Although they are far from the reference field in terms of both static and dynamic data, the inversion results reveal that GTTI spots the similar distribution to the reference field. This is one of the advantages of GTTI – due to quasilinearity, the objective function is less likely to

be stuck at the local minimum even if an initial model is quite different from the reference field.

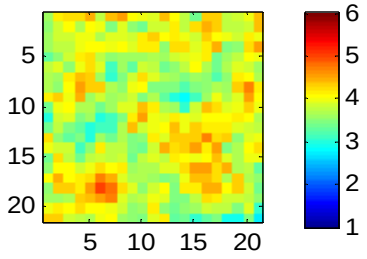
The 3rd representative model shows similar distribution to the reference field. It is the model that is clustered with the reference model when the reference model is added in clustering. Therefore, the inversion is completed without much changes compared to other models.



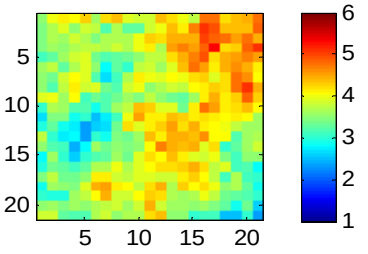
(a)-1 Model 1, initial



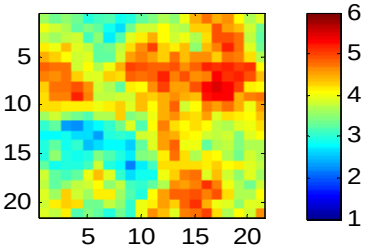
(a)-2 Model 1, matched



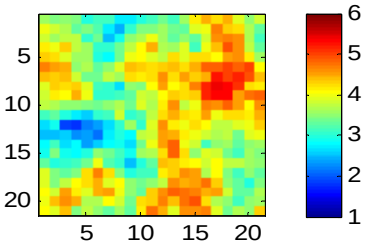
(b)-1 Model 2, initial



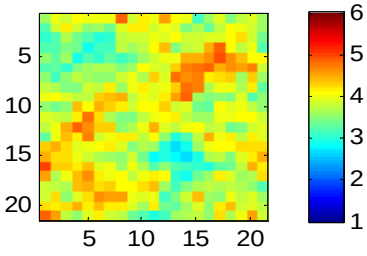
(b)-2 Model 2, matched



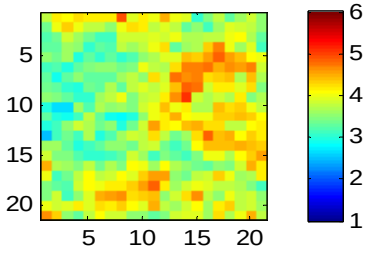
(c)-1 Model 3, initial



(c)-2 Model 3, matched

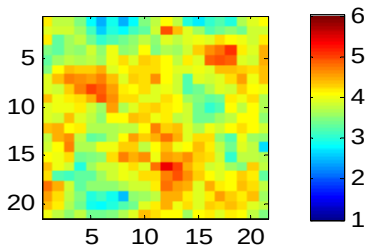


(d)-1 Model 4, initial

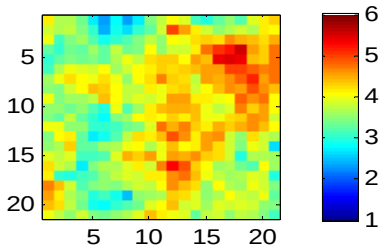


(d)-2 Model 4, matched

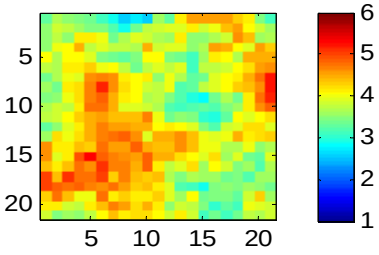
Figure 17 Updated log permeability of representative models (Models 1 to 4)



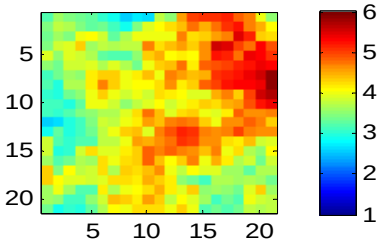
(a)-1 Model 5, initial



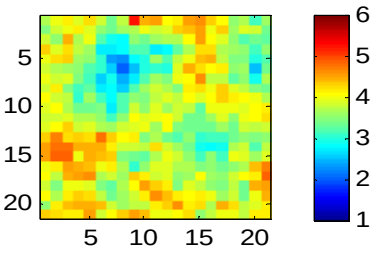
(a)-2 Model 5, matched



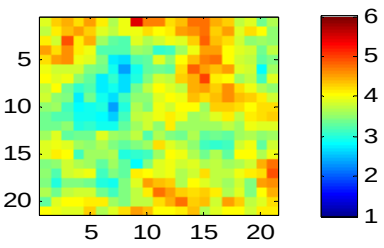
(b)-1 Model 6, initial



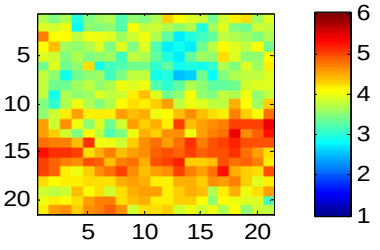
(b)-2 Model 6, matched



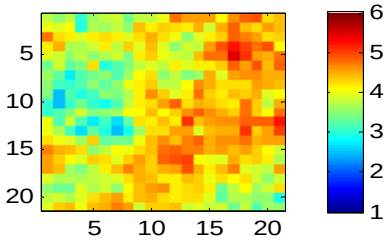
(c)-1 Model 7, initial



(c)-2 Model 7, matched

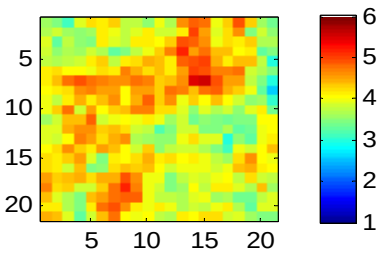


(d)-1 Model 8, initial

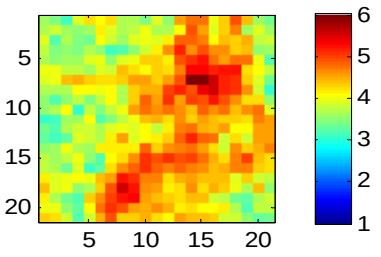


(d)-2 Model 8, matched

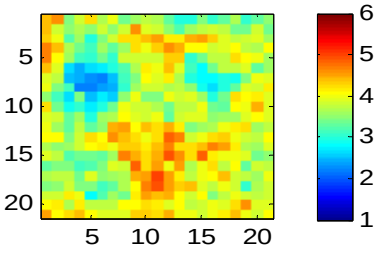
Figure 18 Updated log permeability of representative models (Models 5 to 8)



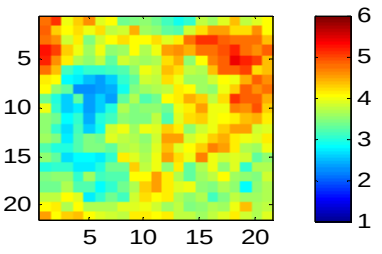
(a)-1 Model 9, initial



(a)-2 Model 9, matched



(b)-1 Model 10, initial



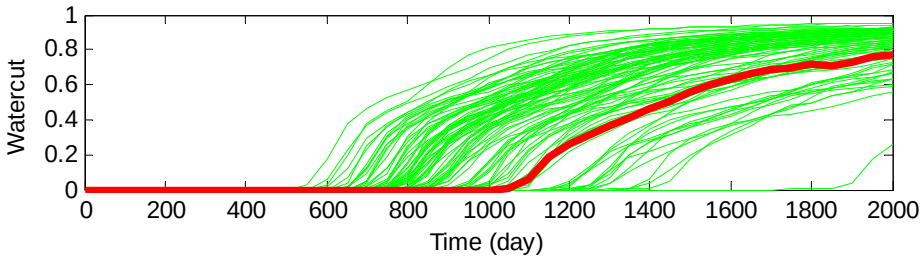
(b)-2 Model 10, matched

Figure 19 Updated log permeability of representative models (Models 9,10)

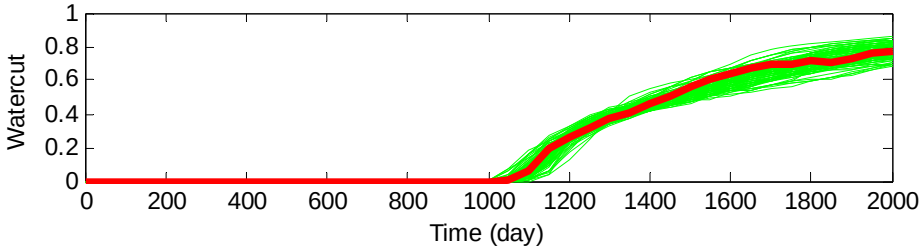
Figures 20 to 27 display initial responses and history matched results from RML. The thick line is the watercut response from the reference model, and the thin lines are responses from each realization. Plots (a) are the results that used all models in the inversion, which Cheng (2003) proposed. Plots (b) are the results of the proposed method. Notation 1 means the initial responses, and 2 means the history matched results. They are grouped by wells for comparison.

As seen in Figures 20 to 27, the suggested method selects models according to the ‘density’ of watercut curves. In other words, they are extracted in unbiased manner from initial one hundred models. In both of the methods, the thick curve is located within the band of the thin curves, which means that the results of inversions capture the dynamic response of the reference field. It is observed that the uncertainty ranges of both of the methods are similar. It indicates that with a few selected models, it is possible to assess the uncertainty reasonably.

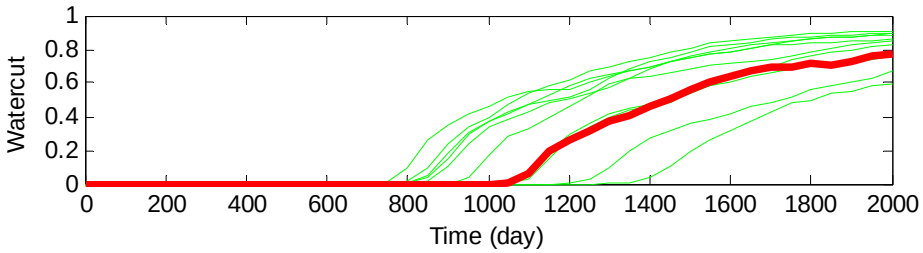
Travel times that watercut reaches the value of 0.5 are plotted in Figures 28 to 35 by wells with boxplots for uncertainty quantification. Each travel time is divided by the travel time of the references field for regularization. Therefore, if the boxplot contains the value 1, we can say that the inversion results capture the true travel time. The result also shows that the proposed method quantifies the uncertainty effectively with a few representative models selected.



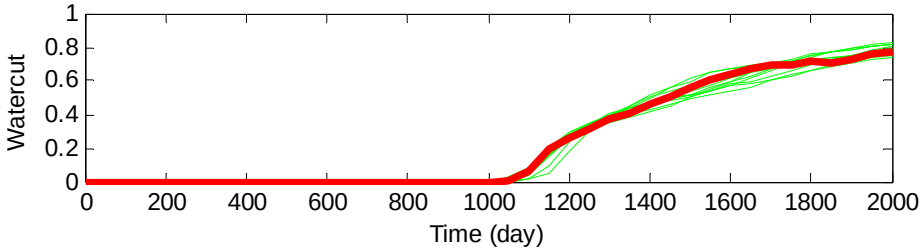
(a)-1 Initial responses from 100 models



(a)-2 History matched result from 100 models

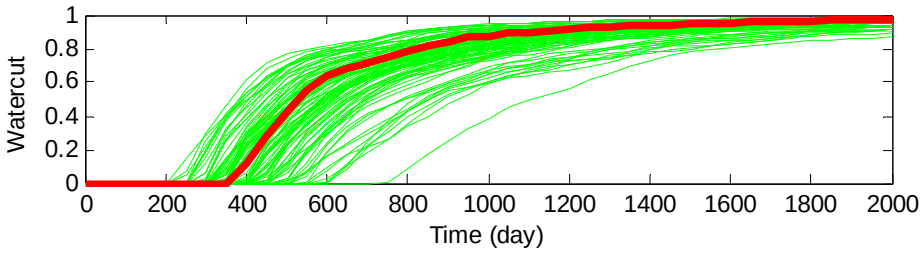


(b)-1 Initial responses from 10 representative models

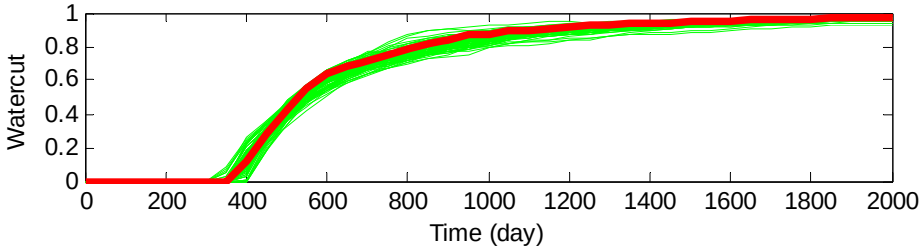


(b)-2 History matched result from 10 representative models

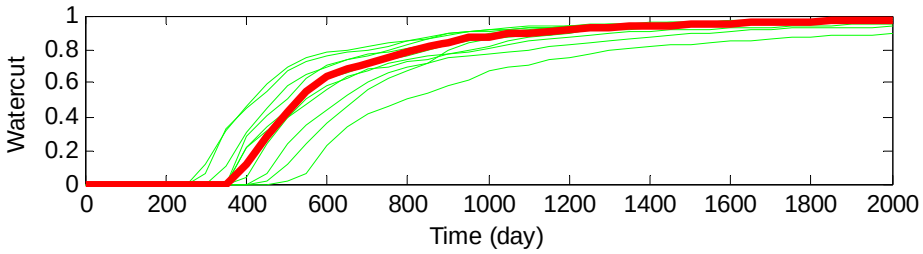
Figure 20 Application of RML on P1 (a) using all models
(b) using representative models



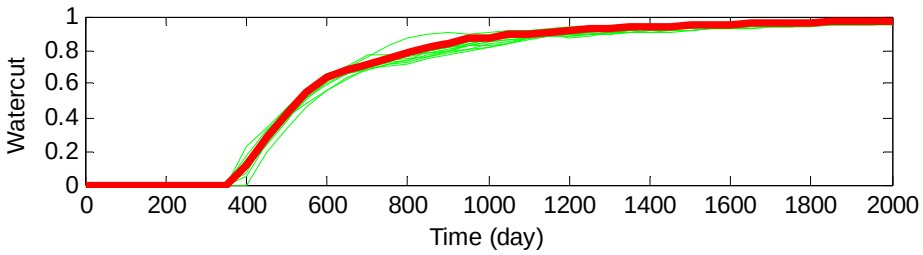
(a)-1 Initial responses from 100 models



(a)-2 History matched result from 100 models

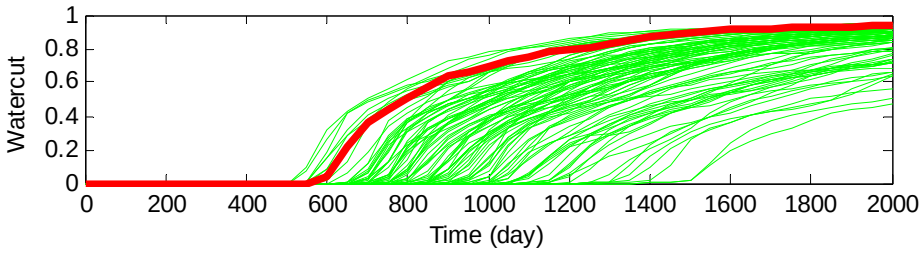


(b)-1 Initial responses from 10 representative models

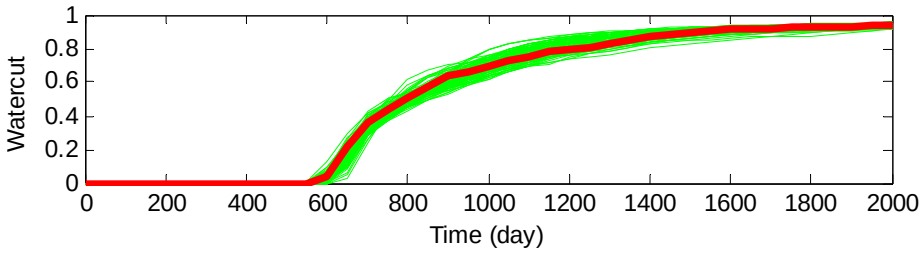


(b)-2 History matched result from 10 representative models

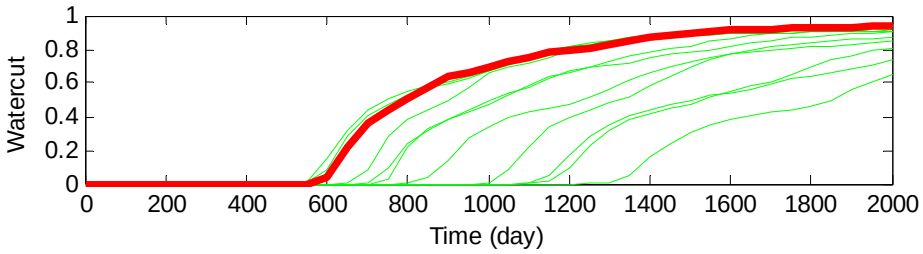
Figure 21 Application of RML on P2 (a) using all models
(b) using representative models



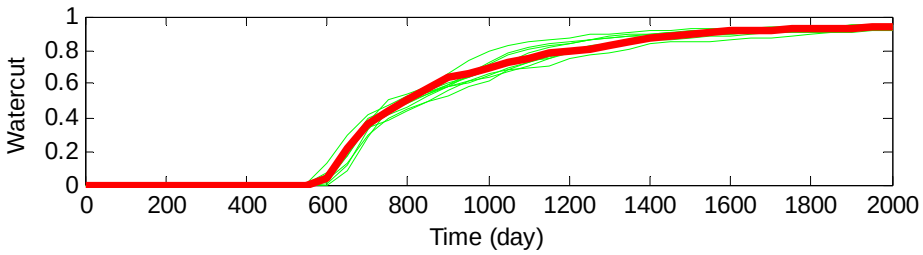
(a)-1 Initial responses from 100 models



(a)-2 History matched result from 100 models

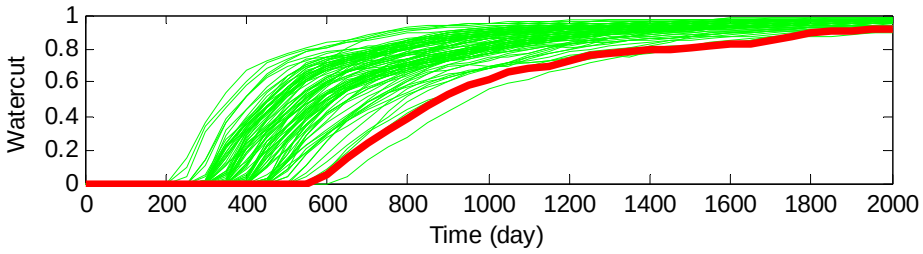


(b)-1 Initial responses from 10 representative models

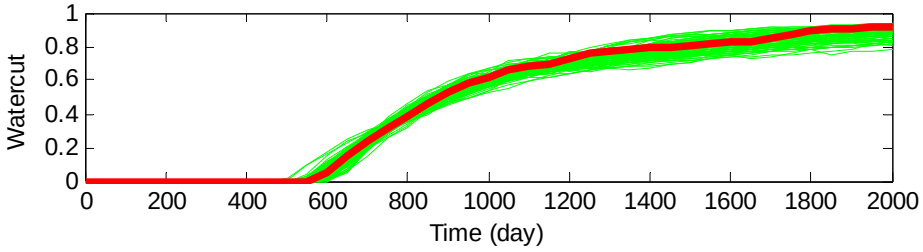


(b)-2 History matched result from 10 representative models

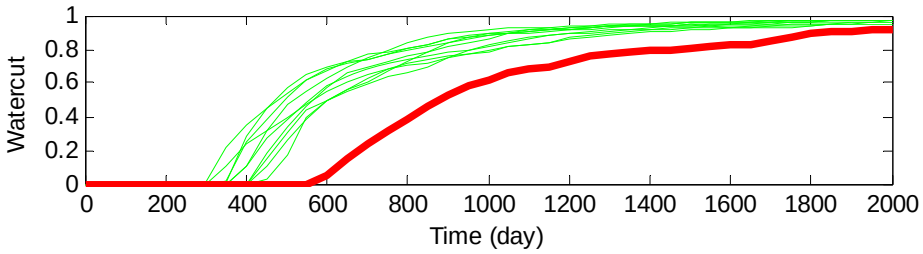
Figure 22 Application of RML on P3 (a) using all models
(b) using representative models



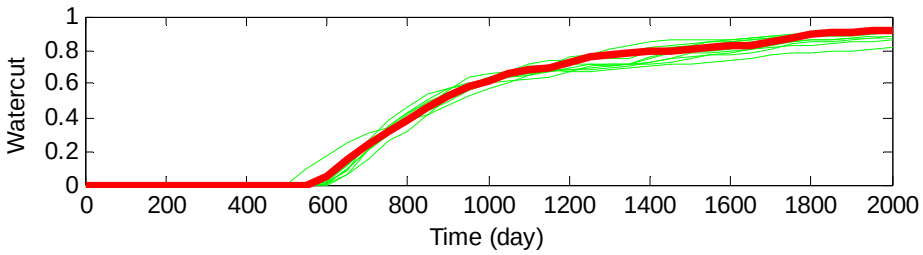
(a)-1 Initial responses from 100 models



(a)-2 History matched result from 100 models

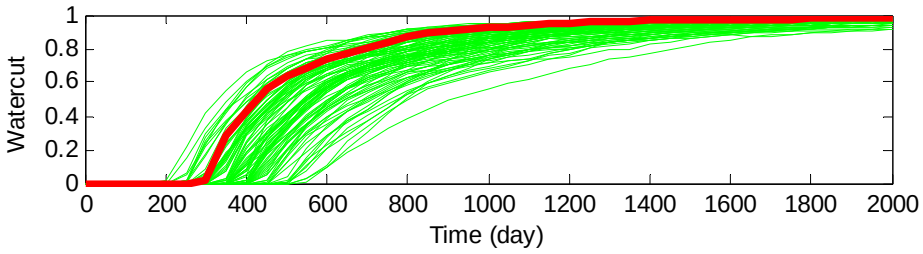


(b)-1 Initial responses from 10 representative models

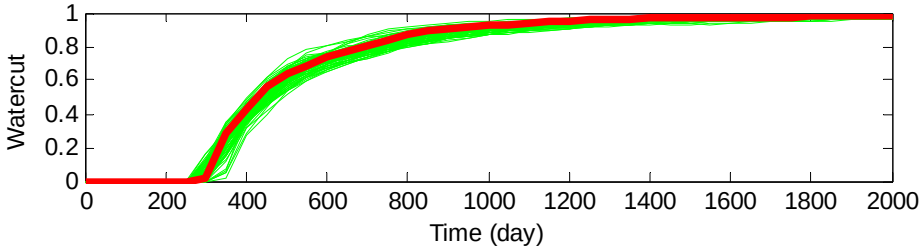


(b)-2 History matched result from 10 representative models

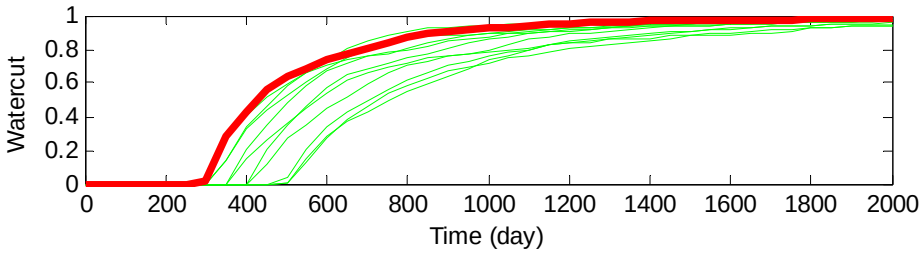
Figure 23 Application of RML on P4 (a) using all models
(b) using representative models



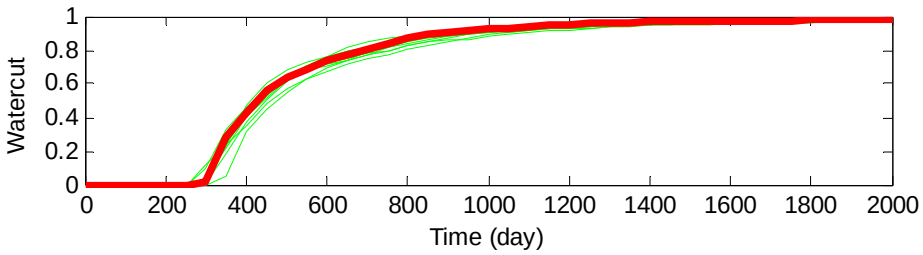
(a)-1 Initial responses from 100 models



(a)-2 History matched result from 100 models

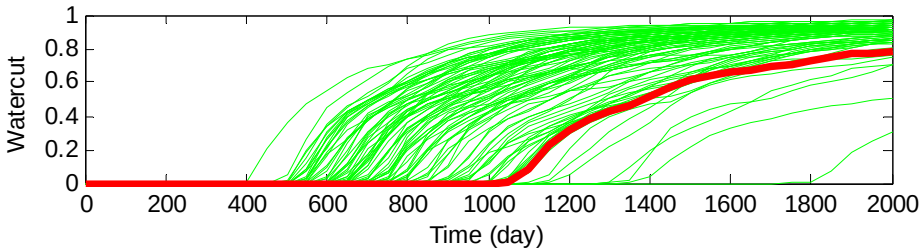


(b)-1 Initial responses from 10 representative models

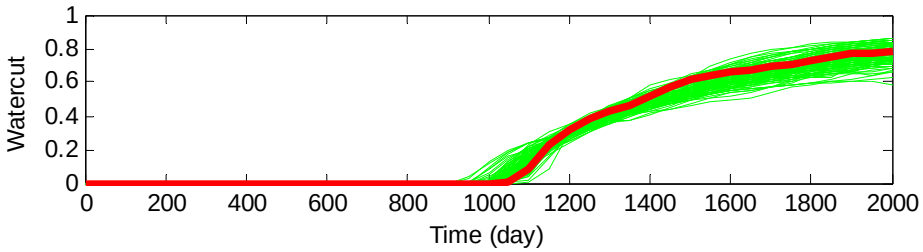


(b)-2 History matched result from 10 representative models

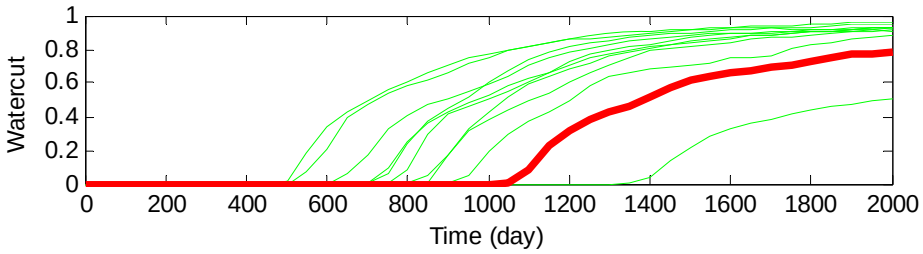
Figure 24 Application of RML on P5 (a) using all models
(b) using representative models



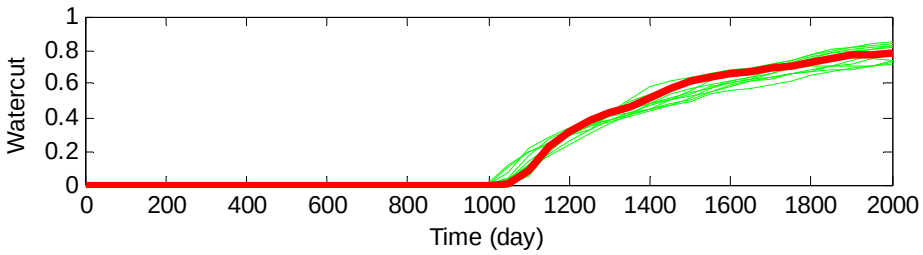
(a)-1 Initial responses from 100 models



(a)-2 History matched result from 100 models

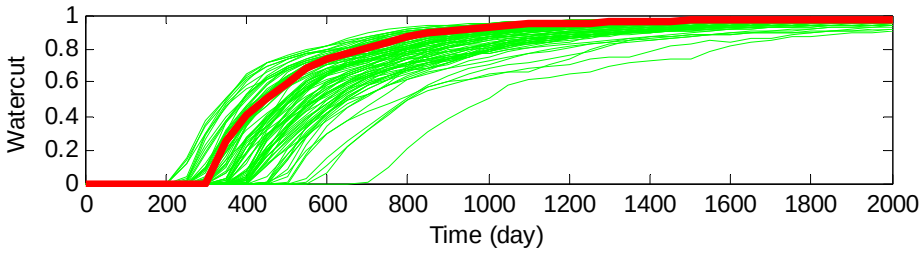


(b)-1 Initial responses from 10 representative models

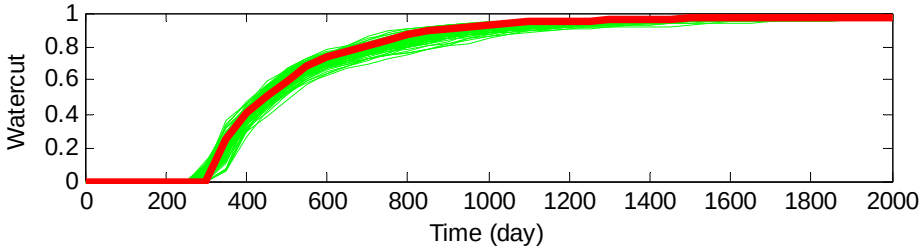


(b)-2 History matched result from 10 representative models

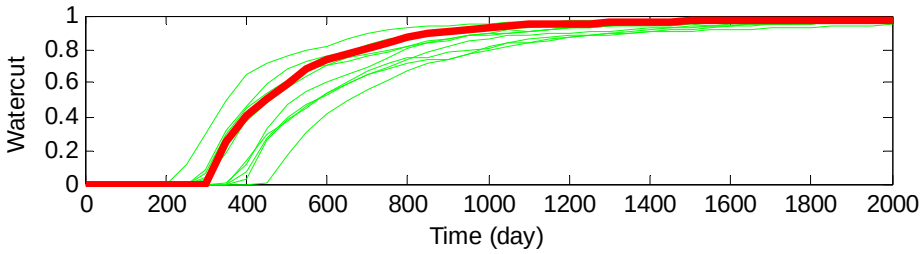
Figure 25 Application of RML on P6 (a) using all models
(b) using representative models



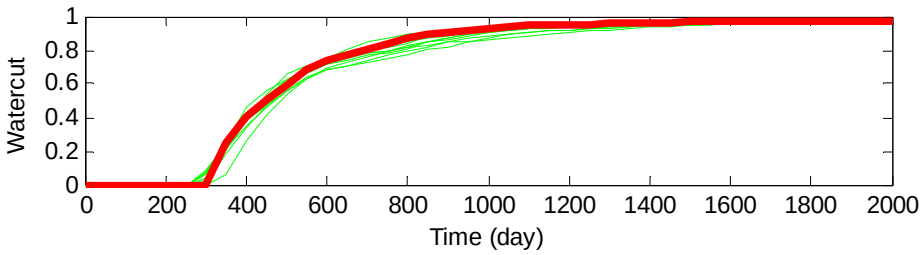
(a)-1 Initial responses from 100 models



(a)-2 History matched result from 100 models

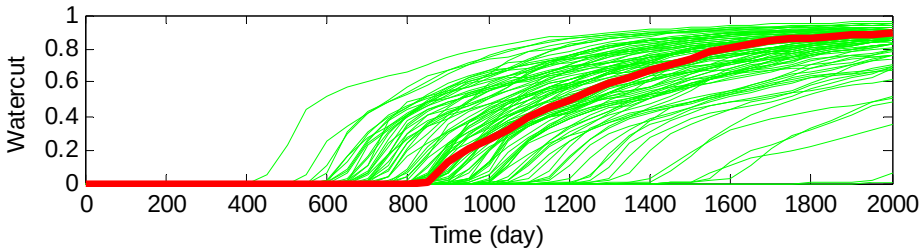


(b)-1 Initial responses from 10 representative models

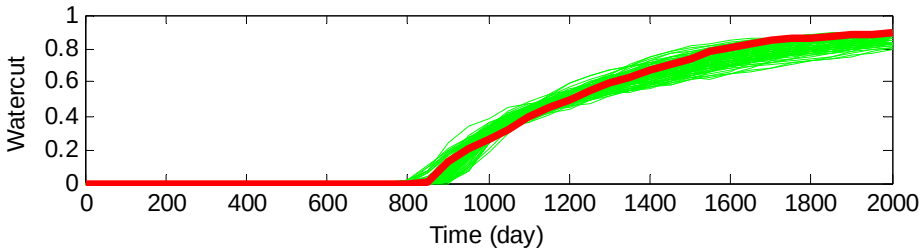


(b)-2 History matched result from 10 representative models

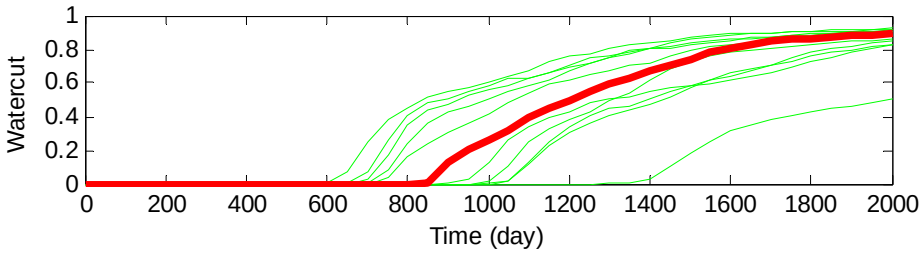
Figure 26 Application of RML on P7 (a) using all models
(b) using representative models



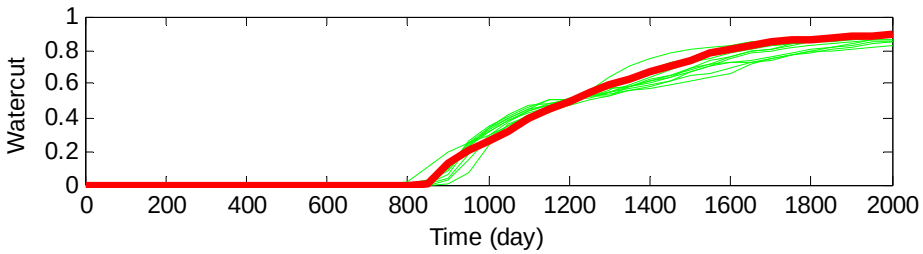
(a)-1 Initial responses from 100 models



(a)-2 History matched result from 100 models



(b)-1 Initial responses from 10 representative models



(b)-2 History matched result from 10 representative models

Figure 27 Application of RML on P8 (a) using all models
(b) using representative models

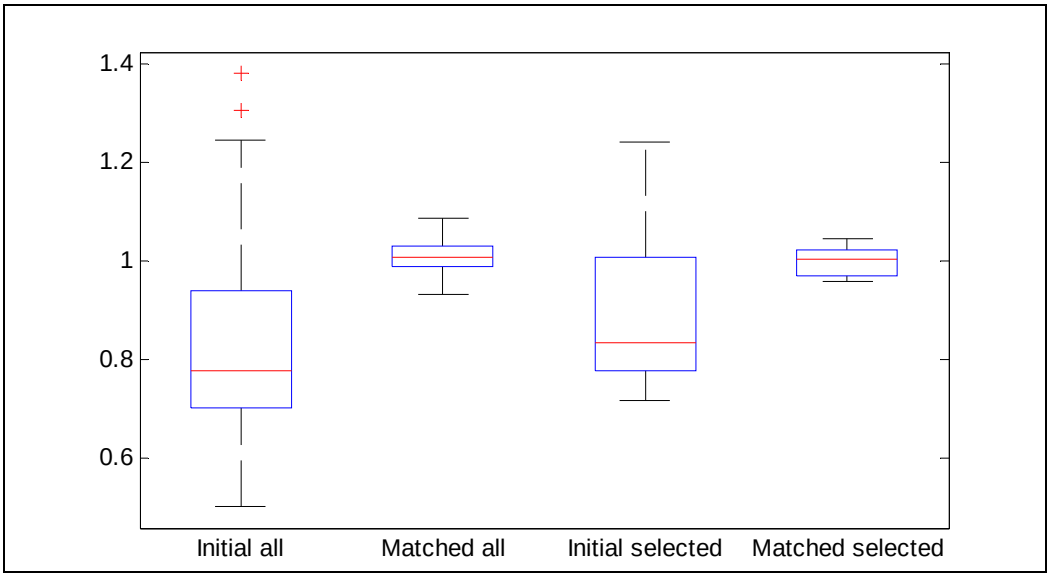


Figure 28 Uncertainty quantification of travel time at P1 (watercut=0.5)

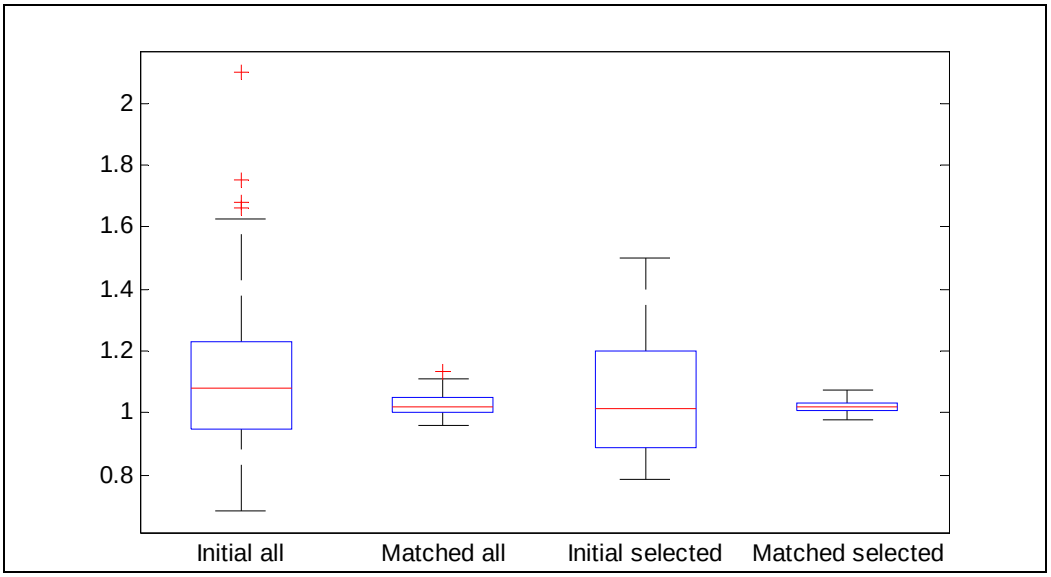


Figure 29 Uncertainty quantification of travel time at P2 (watercut=0.5)

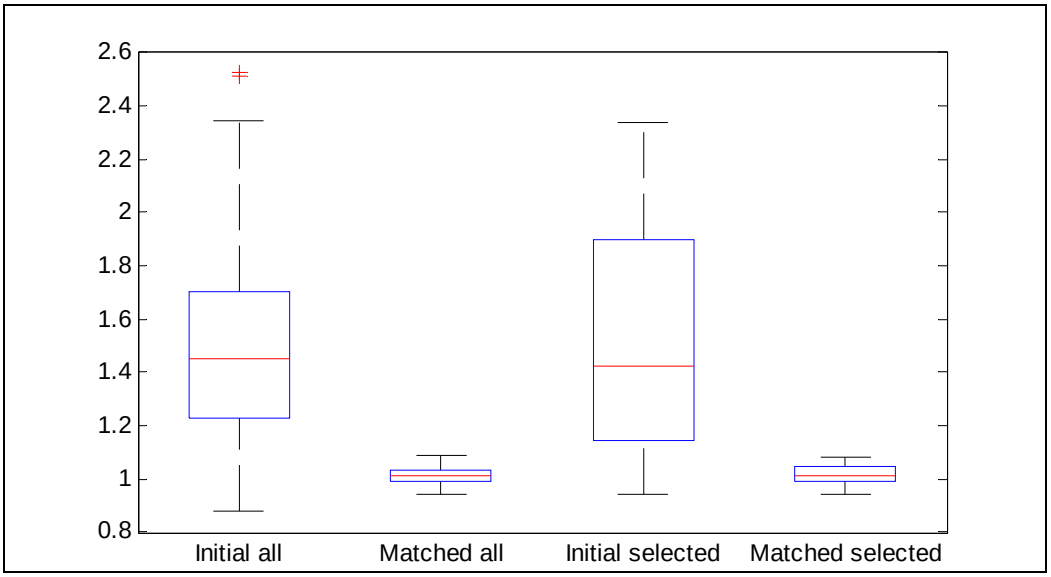


Figure 30 Uncertainty quantification of travel time at P3 (watercut=0.5)

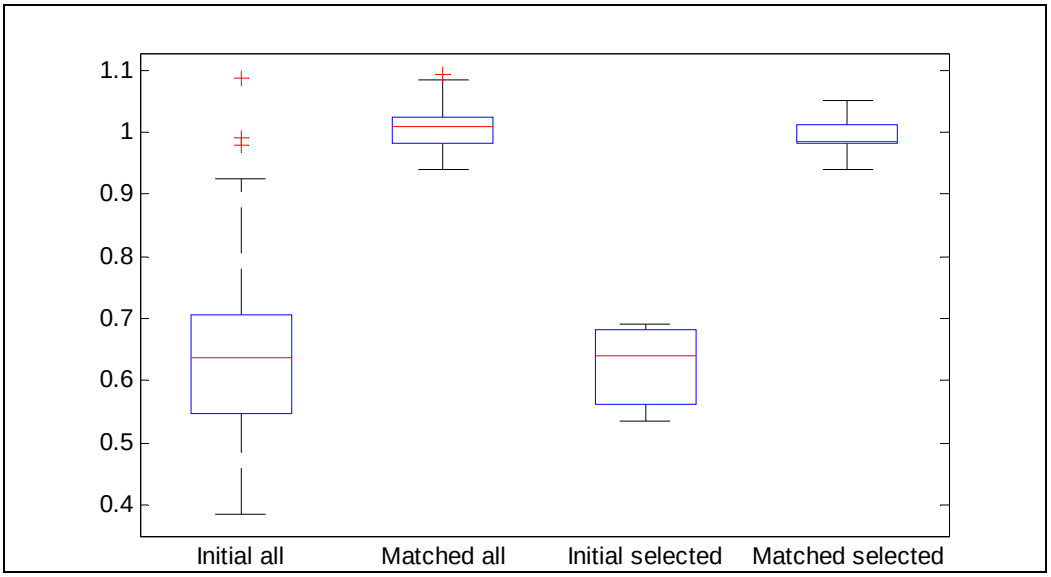


Figure 31 Uncertainty quantification of travel time at P4 (watercut=0.5)

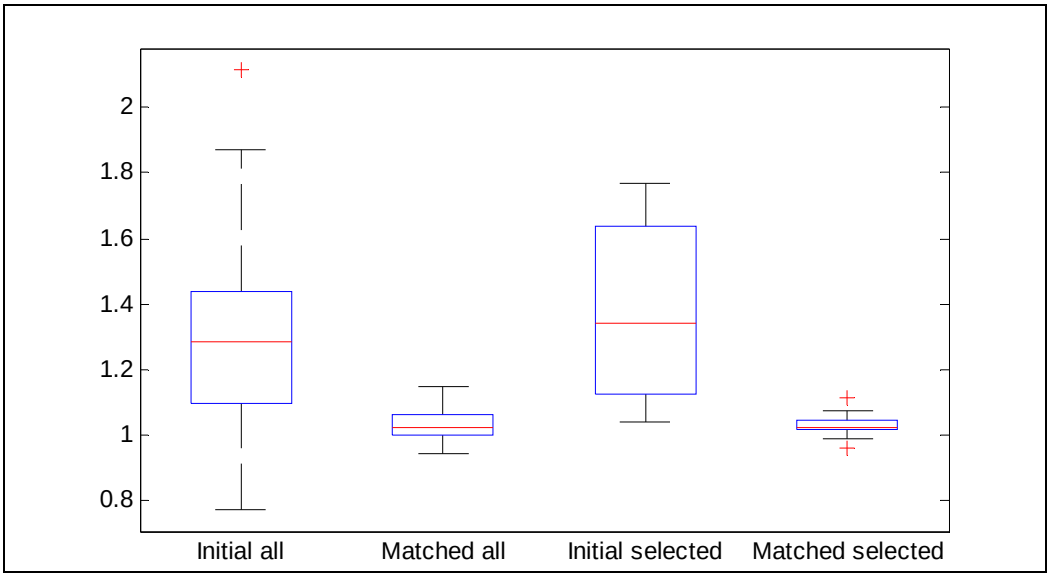


Figure 32 Uncertainty quantification of travel time at P5 (watercut=0.5)

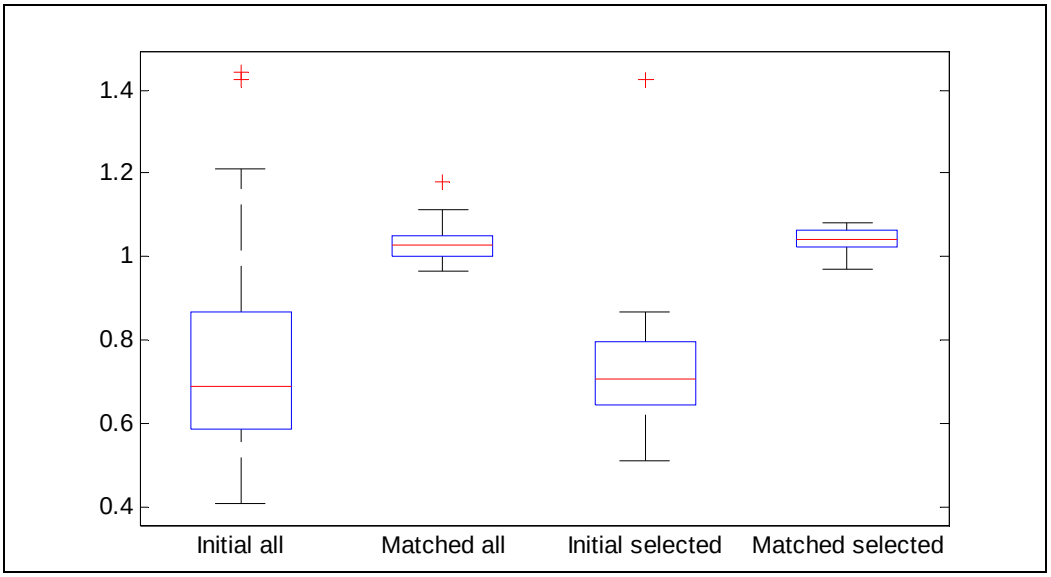


Figure 33 Uncertainty quantification of travel time at P6 (watercut=0.5)

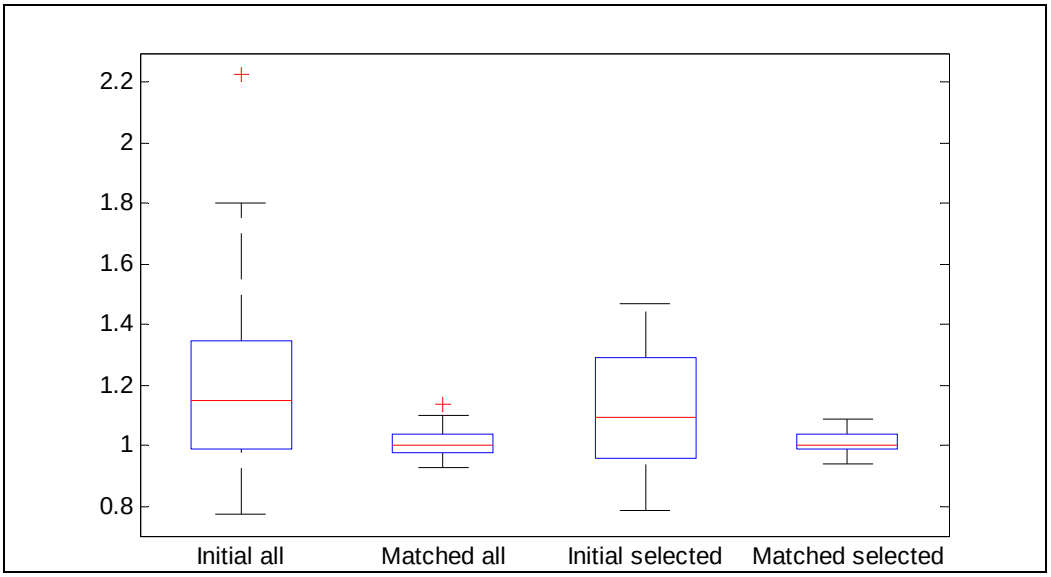


Figure 34 Uncertainty quantification of travel time at P7 (watercut=0.5)

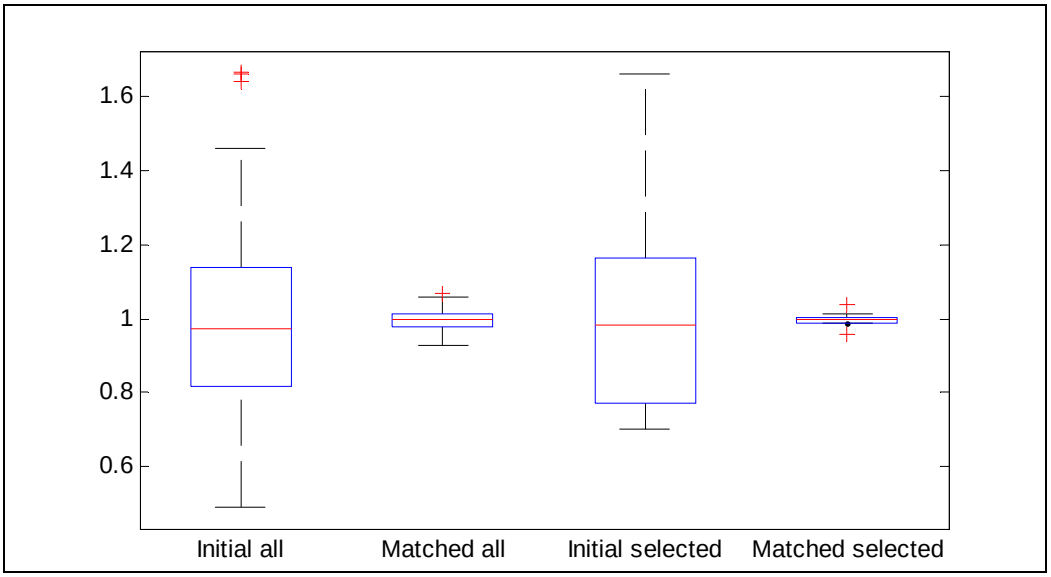


Figure 35 Uncertainty quantification of travel time at P8 (watercut=0.5)

The results of cluster analyses on the travel time by well are displayed in Figures 36 to 43. It is observed that models within the same cluster have similar travel time. Production wells P1, P3, P6, and P8 show comparatively good classification of dynamic data. This is because those wells have wide ranges of travel time in flow responses from initial models.

The suggested clustering method gathers models with similar permeability distribution as well as dynamic data. This is illustrated in Figures 44 to 46 where three models from each cluster are displayed. This is because the distance is defined as GTT differences, and they have high correlations with permeability differences.

To illustrate, Figure 36 indicates the travel time of P1. The models that have the last travel times are gathered in cluster 10, and the earliest travel time in cluster 9. This can be explained by the permeability distribution as plotted in Figure 45b. In Figure 45b, models in cluster 10 have high permeable area on the right bottom, and low area on the left top, where the water sweeps from the injection well to P1. Therefore, we can anticipate that there would be the most delayed breakthrough at cluster 10, which corresponds to the simulation results.

The identical analyses can be applied to other wells. In Figure 38, at P3, models with the earliest breakthrough are gathered in clusters 3 and 4 and models with last breakthrough are gathered in cluster 8. As seen in Figure 44b, cluster 3 has the models that are similar to the reference field. Due to the high permeability streak from the upper right to the bottom, it is expected that there will be early breakthrough at P3. In cluster 4, there are high permeable regions on the upper right, so there will be also early breakthrough at P3. In contrast, we can anticipate that there will be late breakthrough at

P3 for cluster 8, because there is horizontal high permeability streak on the bottom of the models, as seen in Figure 45b. This also matches with the result of Figure 38.

In Figure 41, it is seen that models with the last breakthrough are gathered at cluster 3 for P6. This is also explained by the permeability distribution of each cluster in Figure 44c. Again, in cluster 3, there are low permeability region on the left and high permeability streak on the right. Therefore, much less water sweeps the region of the low permeability area, where streamlines connect between the injection well and P6. This analysis is evident for P8. In Figure 43, cluster 1 gathers the models with the most delayed breakthrough, and as seen in Figure 44a, cluster 1 is characterized as the low permeability area on the right bottom.

To sum up, it is verified that the suggested method effectively gathers models with similar static data as well as dynamic responses. This is because GTT differences reveal the differences in permeability, and they are defined as the distance between models. Furthermore, the validity of the proposed distance based approach is proved, since the dynamic data from each cluster can be analyzed with permeability distribution of models in each cluster in coherent manner.

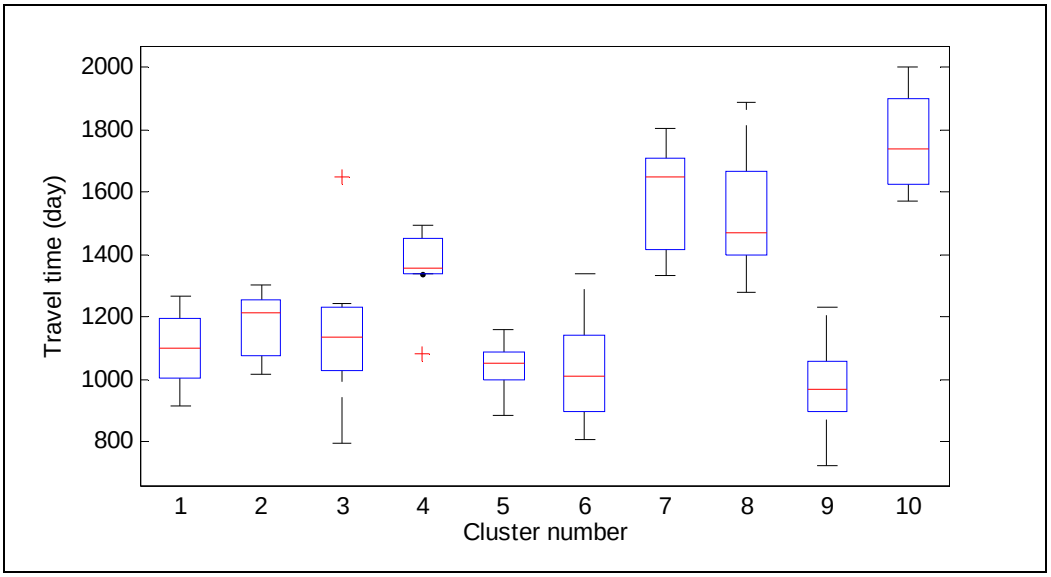


Figure 36 Travel time by cluster at P1 (watercut=0.5)

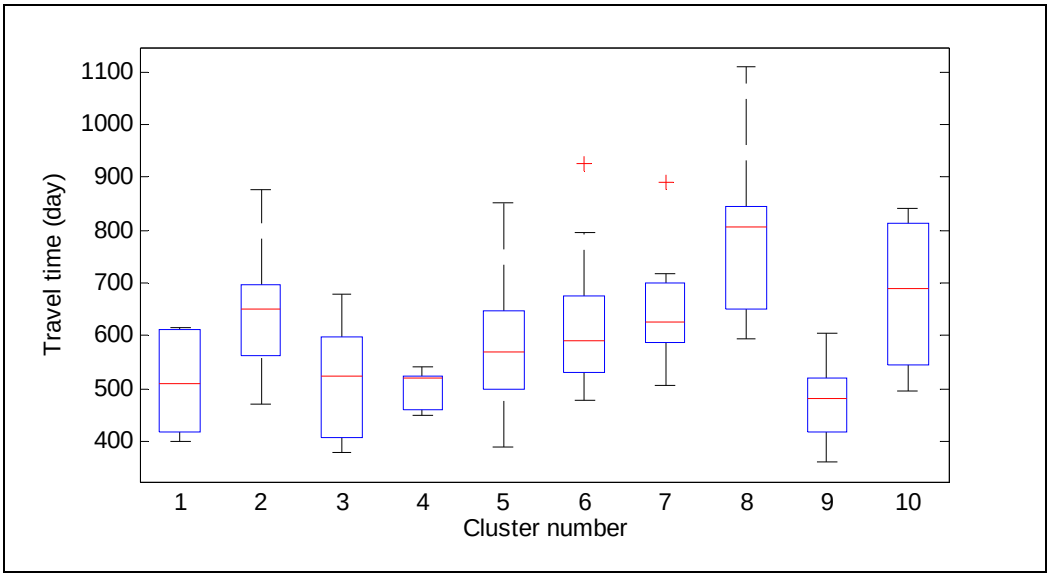


Figure 37 Travel time by cluster at P2 (watercut=0.5)

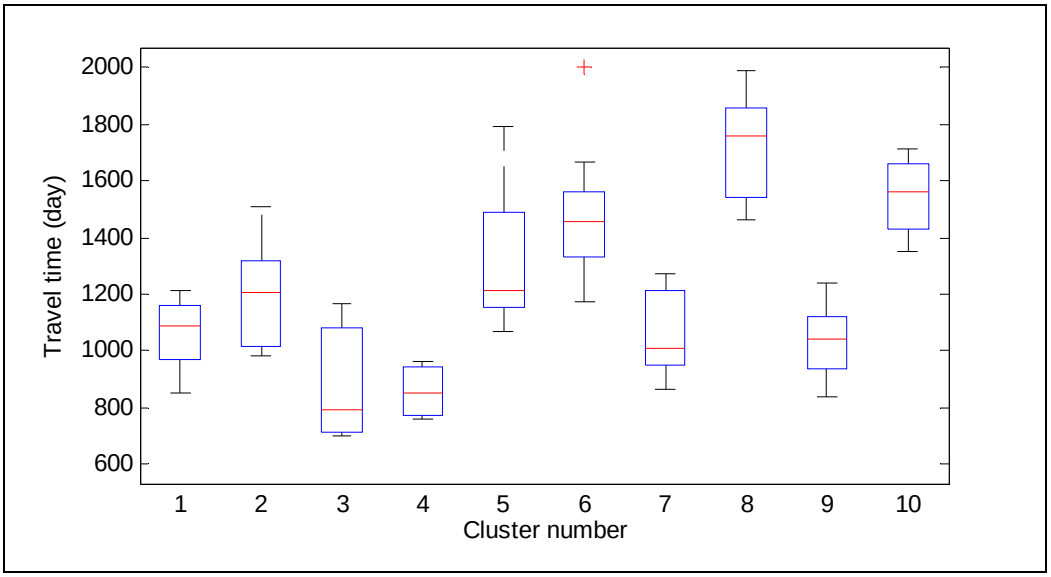


Figure 38 Travel time by cluster at P3 (watercut=0.5)

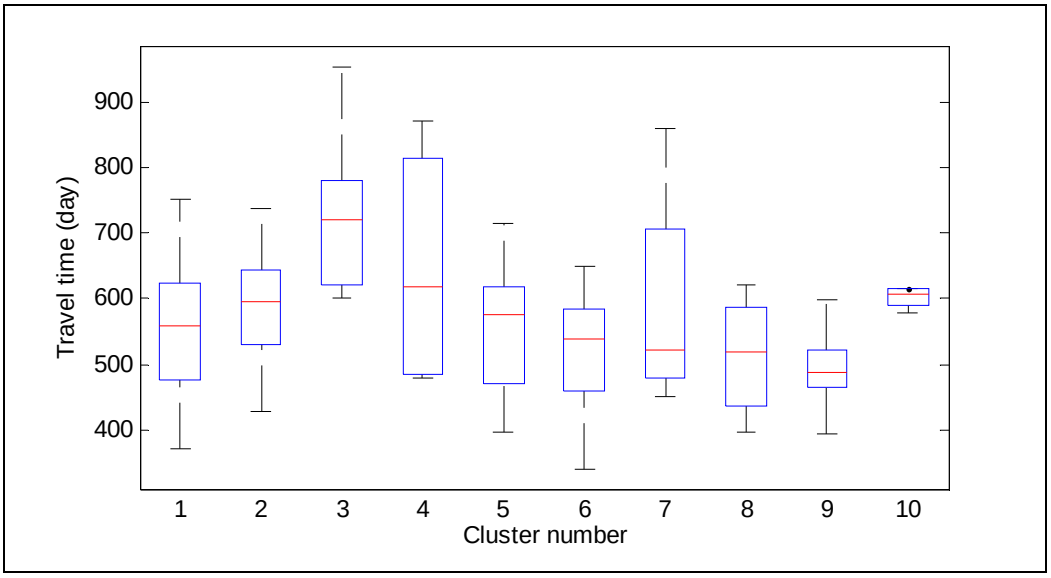


Figure 39 Travel time by cluster at P4 (watercut=0.5)

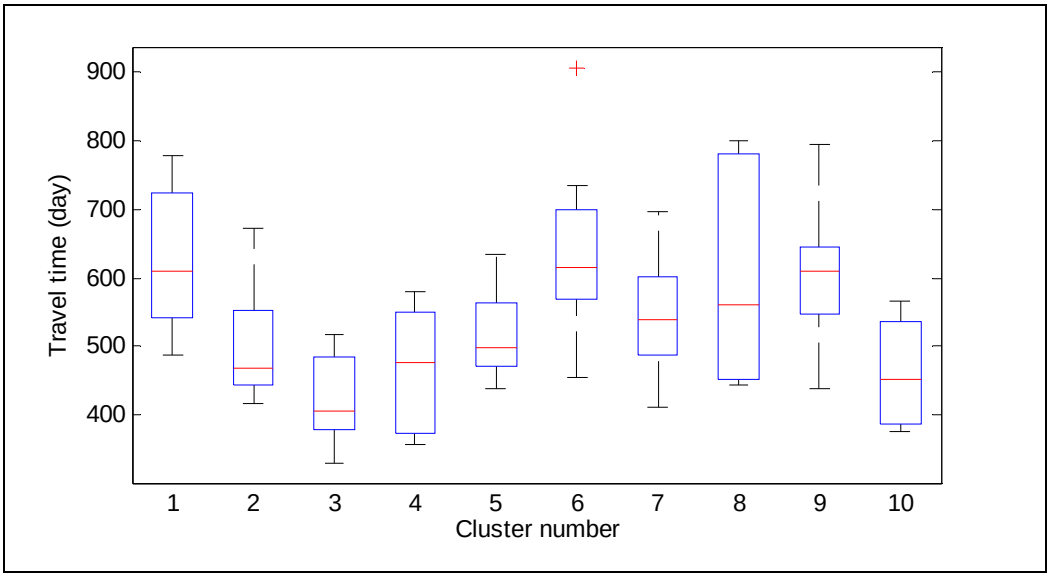


Figure 40 Travel time by cluster at P5 (watercut=0.5)

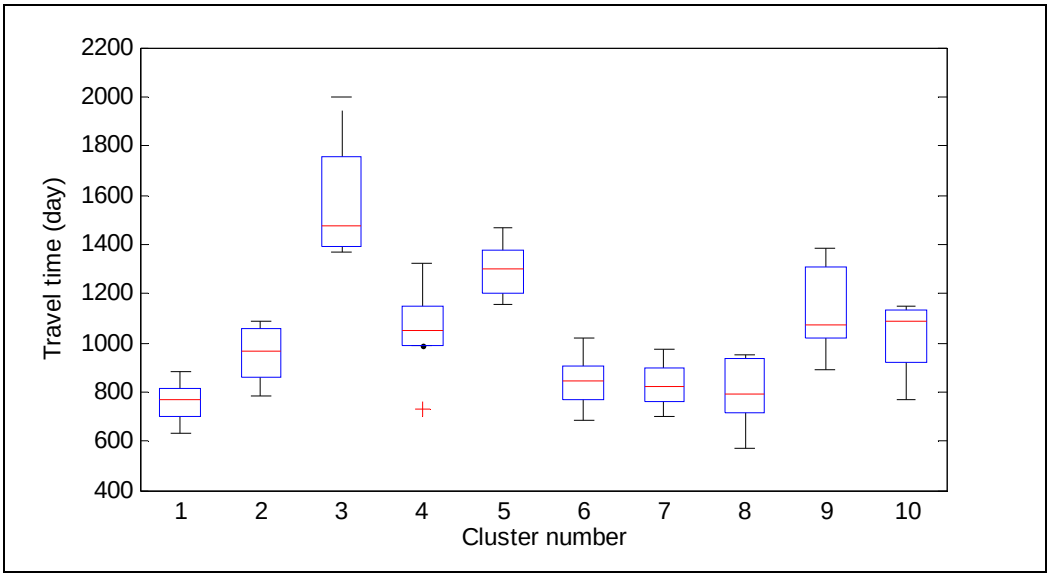


Figure 41 Travel time by cluster at P6 (watercut=0.5)

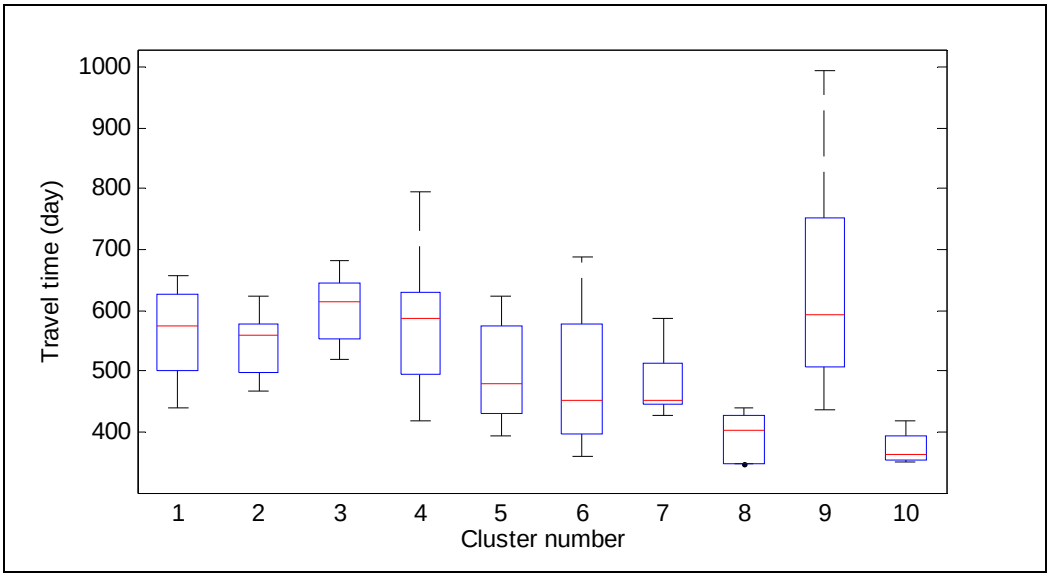


Figure 42 Travel time by cluster at P7 (watercut=0.5)

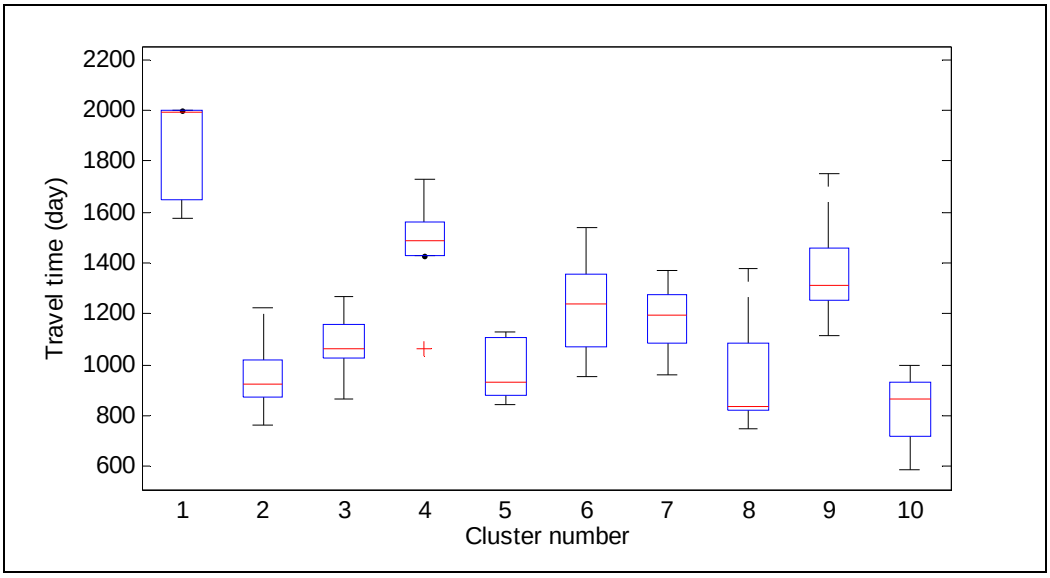
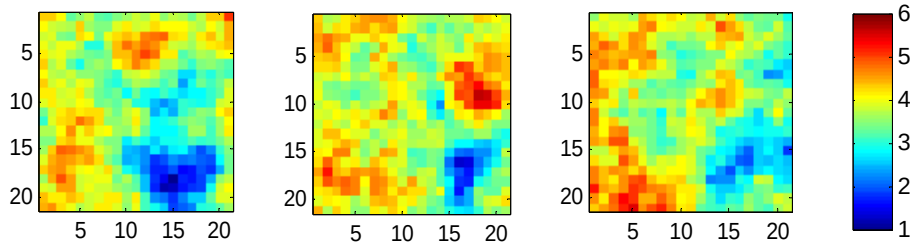
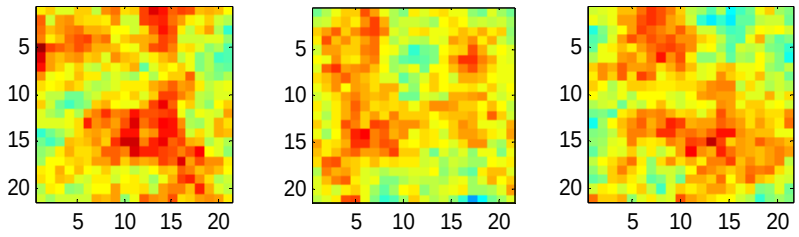


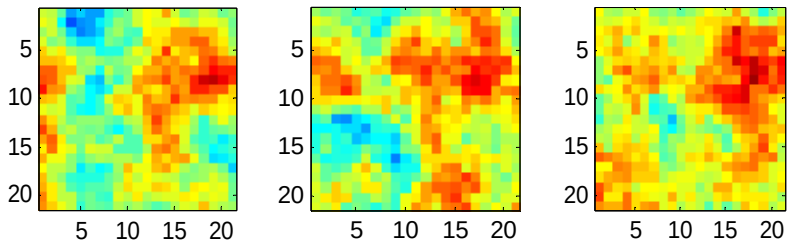
Figure 43 Travel time by cluster at P8 (watercut=0.5)



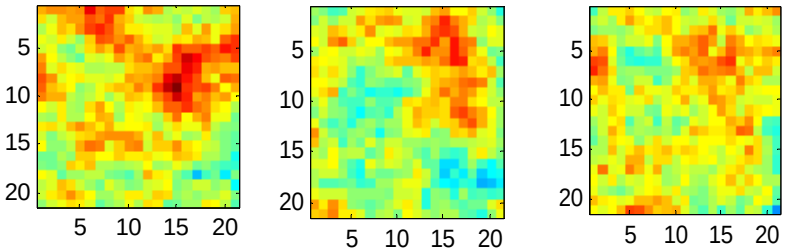
(a) Cluster 1



(b) Cluster 2

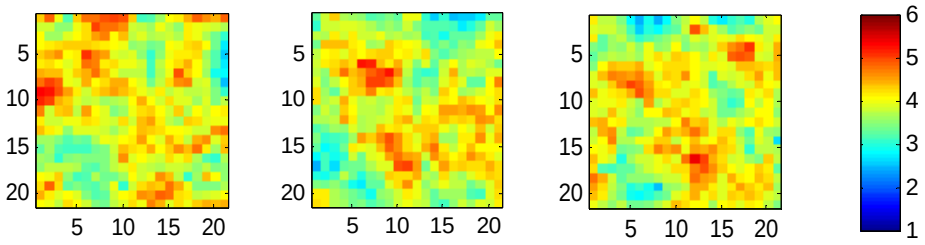


(c) Cluster 3

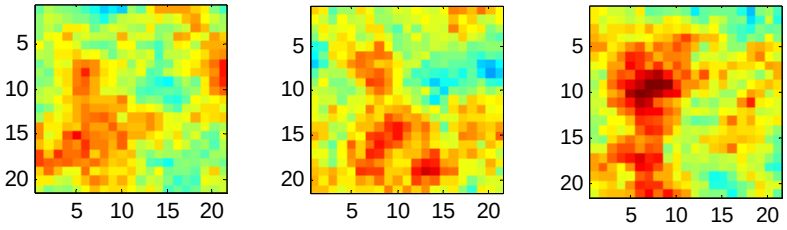


(d) Cluster 4

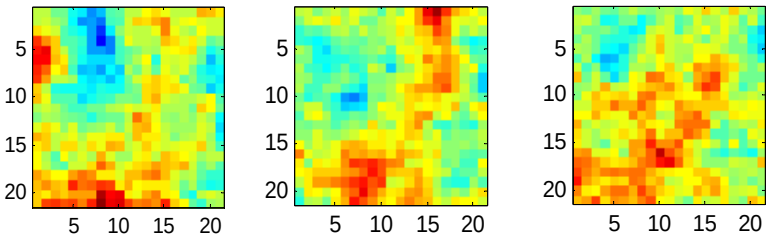
Figure 44 Models from each cluster (clusters 1 to 4)



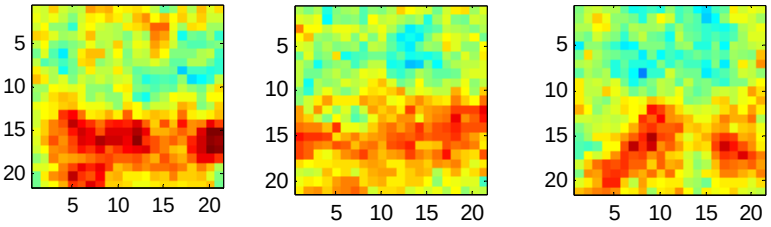
(a) Cluster 5



(b) Cluster 6

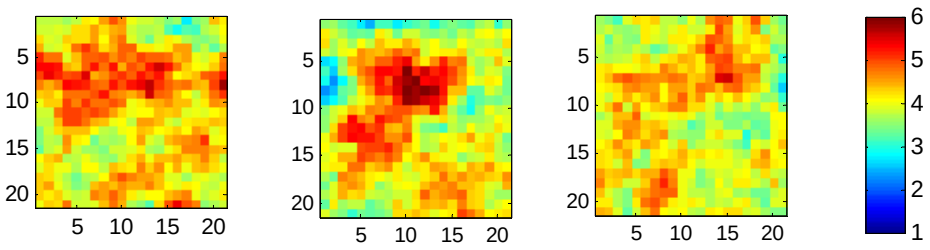


(c) Cluster 7

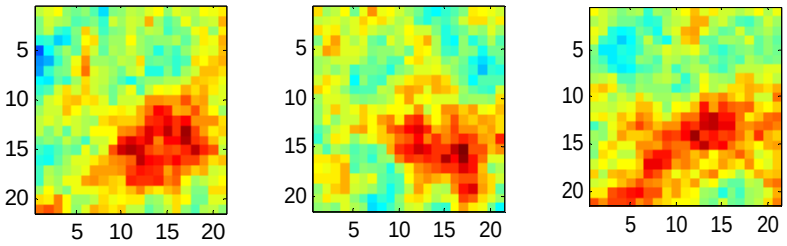


(d) Cluster 8

Figure 45 Models from each cluster (clusters 5 to 8)



(a) Cluster 9



(a) Cluster 10

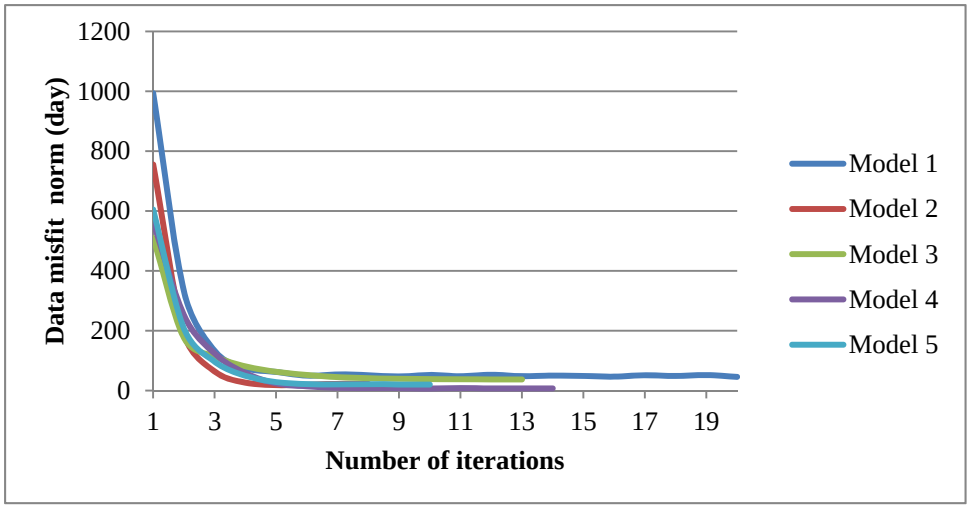
Figure 46 Models from each cluster (clusters 9 and 10)

4.4 Misfit reduction and improvement of computational efficiency

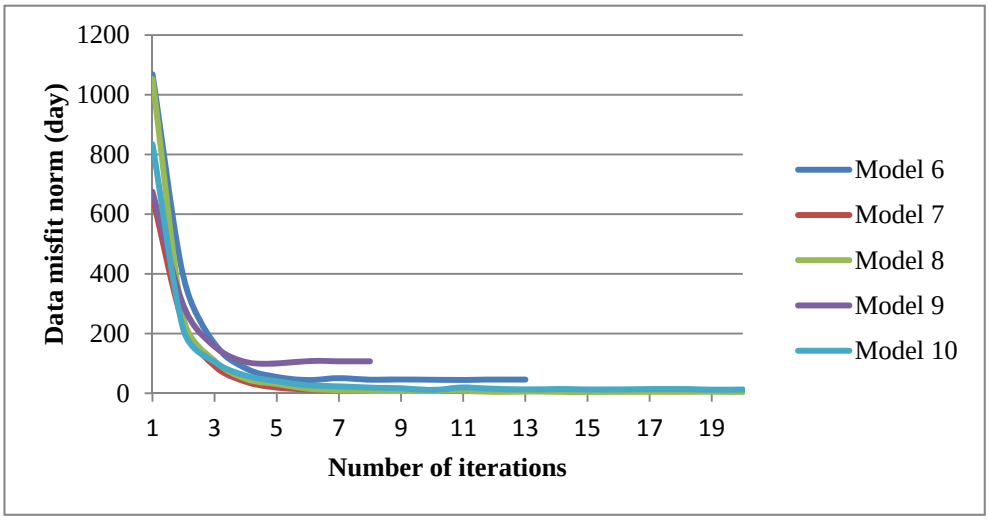
The norms of data misfit with respect to the number of iterations are plotted in Figure 47. The data misfit decreases rapidly at early iterations, and the graphs become stable within about six iterations. It is verified that the objective function converges quickly within a few iterations in GTTI.

In Figure 44, model 3, which is the most similar to the reference field, has the smallest norm value. Model 4, which is also similar to the reference field, also shows the small value. In contrast, Model 7 and Model 8 show large data misfit over 1,000. This is because they have high permeable streak on the reservoir at totally different location compared to the reference field.

Figure 48 compares the number of forward simulations required to perform uncertainty assessment in the two methods – conventional RML that matches all models, and the proposed method that selects representative models after calculating distances and performs history matching. In conventional RML, 1,508 simulations are needed. However, in the proposed method, 246 simulations are needed, which means that the number of forward simulations is reduced by 84.7%. In the proposed method, the forward simulations are comprised of two components – initial simulations for each model to obtain GTT vectors to apply distance based approach, and forward simulations in iteration at the inversion stage. The amount of computations required in clustering is negligible compared to that of the inversion. Therefore, it is concluded that the proposed method significantly enhances the computational efficiency, while reliably assessing the uncertainty.



(a) Data misfit reduction by iterations (Model 1 to 5)



(b) Data misfit reduction by iterations (Model 6 to 10)

Figure 47 Data misfit reduction of each representative model

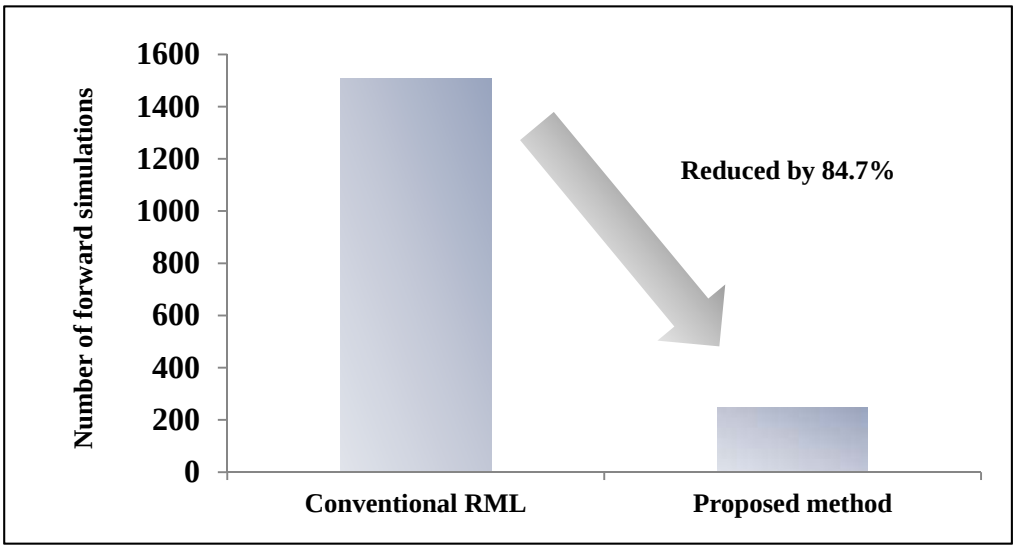


Figure 48 Reduction in the number of forward simulations

5. Conclusions

In the research, a method is proposed to quantify uncertainties with streamline based inversion and distance based method. The research aims at developing a methodology to assess the uncertainty reliably with improved computational efficiency. The distance is defined as the differences of GTT vectors. After applying k-means clustering to group similar models, representative models are selected. Then RML with GTTI is performed on those selected models instead of matching all models. By using streamline simulator as a forward simulator, it is possible to run each simulation rapidly. From the study, the following conclusions can be drawn.

1) A new way of defining a distance is proposed. The distance in the research is defined as the norm of differences of GTT vectors. Defining distance this way, it is possible to group models with similar travel time and permeability distributions. This comes from the fact that GTT is highly dependent on permeability differences.

2) Streamline based inversion is combined with the distance based approach. Before performing RML, models are grouped by k-means clustering using the distance defined in this study. Inversions are performed on the representative models that are selected from each cluster. It is demonstrated that the proposed method effectively quantifies uncertainties while reducing the number of calculations significantly. Compared to conventional RML, the number of forward simulations is decreased by 84.7%.

3) Since RML with GTTI is used, it is possible to fully utilize the advantages of streamline based inversion. Using a streamline simulator as a forward simulator enhances computational speed of forward simulations. In addition, due to quasilinearity of GTTI, the inversion result is less sensitive to the choice of an initial model and the objective function decreases rapidly within a few iterations.

Streamline simulation and streamline based inversion techniques have been proved as effective methods in uncertainty quantifications due to its computational efficiency and robustness. It is expected that the proposed methodology will strengthen these advantages while maintaining the reliability.

References

Batycky, R. P., Blunt, M. J. and Thiele, M. R. 1997. A 3D Field-scale Streamline-Based Reservoir Simulator. *SPE Reservoir Engineering* **12** (4): 246-254.

Caers, J. 2011. *Modeling Uncertainty in the Earth Sciences*. Chichester, UK: John Wiley & Sons Ltd.

Cheng, H., Datta-Gupta, A. and He, Z. 2003. A Comparison of Travel-time and Amplitude Matching for Field-scale Production Data Integration: Sensitivity, Non-linearity and Practical Implications. Paper SPE 84570 presented at the SPE Annual Technical Conference and Exhibition, Denver, Colorado, 5-8 March.

Cheng, H., Wen, X., Milliken, W. and Datta-Gupta, A. 2004. Field Experiences with Assisted and Automatic History Matching Using Streamline Models. Paper SPE 89857 presented at the SPE Annual Technical Conference and Exhibition, Houston, Texas, 26-29 September.

Cheng, H., Kharghoria, A. and Datta-Gupta, A. 2005. Fast History Matching of Finite-difference Models Using Streamline-derived Sensitivities. *SPE Reservoir Evaluation & Engineering* **8** (5): 426-436.

Cheng, H. 2005. Fast History Matching of Finite-difference Model, Compressible and Three-phase Flow Using Streamline-derived Sensitivities. PhD Dissertation, Texas A&M University, College Station, Texas.

Cheng, H., Oyerinde, D. and Datta-Gupta, A. 2006. Compressible Streamlines and Three-phase History Matching. Paper SPE 99465 presented at the SPE/DOE Symposium on Improved Oil Recovery, Tulsa, Oklahoma, 22-26 April.

Choe, J. 2007. *Geostatistics*. Seoul, South Korea: Sigma Press.

Crane, M. J. and Blunt, M. J. 1999. Streamline-based Simulation of Solute Transport. *Water Resources Research* **35** (10): 3061-3078.

Dake, L. P. 1978. *Fundamentals of Reservoir Engineering.* Amsterdam, Netherlands: Elsevier.

Datta-Gupta, A. and King, M. J. 1995. A Semianalytic Approach to Tracer Flow Modeling in Heterogeneous Permeable Media. *Advances in Water Resources* **18** (1): 9-24.

Datta-Gupta, A. 2000. Streamline Simulation: a Technology Update. *Journal of Petroleum Technology* **52** (12): 68-73.

Datta-Gupta, A. and King, M. J. 2007. *Streamline Simulation: Theory and Practice*. Richardson, Texas: SPE.

Emanuel, A. and Milliken, W. 1998. History Matching Finite Difference Models with 3D Streamlines. Paper SPE 49000 presented at the SPE Annual Technical Conference and Exhibition, New Orleans, Louisiana. 27-30 September.

Ertekin, T., Abou-Kassem, J. H. and King, G. R. 2001. *Basic Applied Reservoir Simulation*. Richardson, Texas: SPE.

Gao, G., Zafari, M. and Reynolds, A. C. 2005. Quantifying Uncertainty for the PUNQ-S3 Problem in a Bayesian Setting with RML and EnKF. Paper SPE 93324 presented at the SPE Reservoir Simulation Symposium, Houston, Texas, 31 January-2 February.

He, Z., Yoon, S. and Datta-Gupta, A. 2002. Streamline-based Production Data Integration with Gravity and Changing Field Conditions. *SPE* Journal **7** (4): 423-436.

Jang, M. and Choe, J. 2001. Modeling of Contaminants Transport in a Fracture Using Streamline Simulation. *Journal of Korean Society of Geosystem Engineering* **38** (3): 202-210.

Jin, J. 2011. Oil Production Analysis in Oil Sands Reservoirs Using Distance Based Clustering. MS Thesis, Seoul National University, Seoul, South Korea.

Jung, H. 2008. Reservoir Characterization Using Gradual Deformation Method with Ensemble Kalman Filter. MS Thesis, Seoul National University, Seoul, South Korea.

Kitanidis, P. K. 1995. Quasilinear Geostatistical Theory for Inversing. *Water Resources Research* **31** (10): 2411-2419.

Milliken, W., Emanuel, A. and Chakravarty, A. 2001. Applications of 3D Streamline

Simulation to Assist History Matching. *SPE Reservoir Evaluation & Engineering* **4** (6): 502-508.

Oliver, D. S., Reynolds, A. C. and Liu, N. 2008. *Inverse Theory for Petroleum Reservoir Characterization and History Matching*. Cambridge, UK: Cambridge University Press.

Oliver, D. S. and Chen, Y. 2011. Recent Progress on Reservoir History Matching: a Review. *Computational Geosciences* **15** (1): 185-221.

Oyerinde, A. 2008. Streamline-based Three-phase History Matching. PhD Dissertation, Texas A&M University, College Station, Texas.

Oyerinde, A., Datta-Gupta, A. and Milliken, W. 2009. Experiences with Streamline-based Three-phase History Matching. *SPE Reservoir Evaluation & Engineering* **12** (4): 528-541.

Park, W. 2010. Efficient Sensitivity Calculation Using Ensemble Method. MS Thesis, Seoul National University, Seoul, South Korea.

Pollock, D. W. 1988. Semianalytical Computation of Path Lines for Finite-difference Models. *Ground Water* **26** (6): 743-750.

Remy, N., Boucher, A. and Wu, J. 2009. *Applied Geostatistics with SGeMS: a User's Guide*. Cambridge, UK: Cambridge University Press.

Sandve, T. H. 2009. Streamline Based History Matching. MS Thesis, University of Bergen, Bergen, Norway.

Scheidt, C. and Caers, J. 2009. Representing Spatial Uncertainty Using Distances and Kernels. *Mathematical Geosciences* **41** (4): 397-419.

Scheidt, C. and Caers, J. 2009. Uncertainty Quantification in Reservoir Performance Using Distances and Kernel methods-Application to a West Africa Deepwater Turbidite Reservoir. *SPE Journal* **14** (4): 680-692.

Thiele, M. R. 2005. Streamline Simulation. Paper presented at *8th International Forum on Reservoir Simulation*, Lago Maggiore, Italy, 20-24 June.

Vasco, D. W., Yoon, S. and Datta-Gupta, A. 1999. Integrating Dynamic Data into High-resolution Reservoir Models Using Streamline-based Analytic Sensitivity Coefficients. *SPE Journal* **4** (4): 389-399.

Vega, L., Rojas, D. and Datta-Gupta, A. 2004. Scalability of the Deterministic and Bayesian Approaches to Production-data Integration into Reservoir Models. *SPE Journal* **9** (3): 330-338.

Willhite, P. G. 1986. *Waterflooding*. Richardson, Texas: SPE.

Wu, Z. and Datta-Gupta, A. 2001. Rapid History Matching Using a Generalized Travel Time Inversion Method. Paper SPE 66352 presented at the *SPE Reservoir*

Simulation Symposium, Houston, Texas, 11-14 February.

Yeo, M. 2010. Reservoir Characterization Using Ensemble Kalman Filter with Drainage Area Localization. MS Thesis, Seoul National University, Seoul, South Korea.

국문초록

저류층특성화 및 저류층 거동의 불확실성 평가는 석유개발사업을 위한 필수요건이다. 제한된 자료로 생성된 초기모델은 큰 불확실성을 가지기 때문에 신뢰할 수 있는 예측을 위해서는 정적, 동적 자료를 모델에 결합해야 한다. 이러한 불확실성 정량화 과정은 하나의 모델에 대한 역산을 위해 반복적인 전위시뮬레이션과 최적화계산이 필요하며 이 과정을 다수의 모델에 대해 실시하여야 하므로 많은 계산량이 요구된다. 또한 히스토리 매칭은 수학적으로 잘 정립되지 못해 역산이 어렵다.

본 논문에서는 유선기반역산과 거리기반방법을 결합하여 빠르게 불확실성을 정량화하는 방법을 제시하였다. 첫째, 모델간의 거리는 generalized travel time 차이의 norm 으로 정의하였다. 둘째, 생성한 모든 모델에 대해 매칭을 실시하는 대신, 거리에 따라 모델을 군집화한 후 대표모델을 선택하여 역산을 실시하였다. 셋째, generalized travel time inversion 알고리즘을 이용하여 동적자료를 결합하였고 전위시뮬레이터로 유선시뮬레이터를 사용하여 계산효율을 향상하였다. 제안된 방법은 비슷한 동적자료와 유체투과율 분포를 가지는 모델을 효과적으로 군집화함이 확인되었다. 또한 대표모델만을 사용하였기 때문에 계산량을 획기적으로 감소시키면서도 불확실성을 잘 평가하였다.

주요어: 유선시뮬레이션, generalized travel time inversion, 거리기반방법, 저류층특성화

학 번: 2012-20999